Tilhoff, Tanya

From: Akerley, Marj
Sent: 2015-May-25 4:52 PM
To: Topshee, Dugald
Cc: Roy, Dominique; Sampson, Tracey; Jakob, David; Poirier, Linda; Mauchan, Lana
Subject: Fw: Request participation in a project about the possibilities of Big Data within the Department
Attachments: Background Big Data and Privacy Project.docx

Importance: High

Dugald,

Can you.pls work with my office to schedule. We should also include David and Dominique as required. Either Tracey or I will attend (I would like to go if possible but will depend on timing).

Thx
Marj

Marj Akerley
Chief Information Officer, Dirigeante Principale de l'information

Sent from my Blackberry

From: Fortin, Suzanne <Suzanne.Fortin@justice.gc.ca>
Sent: Monday, May 25, 2015 2:40 PM
To: Akerley, Marj; Poirier, Linda
Cc: MacLean, Alyson; Fraser, Charlotte; Lipinski, Stan; Di Duca, Pauline
Subject: Request participation in a project about the possibilities of Big Data within the Department

Good afternoon Marj,

I am writing to you on behalf of Stan Lipinski (DG PICS, Policy Sector) for whom I am acting today.

The Research and Statistics Division, Policy Sector has contracted with E.S. Tunis and Associates to investigate emerging trends affecting the current and possible future uses of Big Data (and Privacy) within the Department. This work is supporting direction from our Deputy Minister to engage in forward-looking exercises to explore issues that may impact the Department in the future. The first phase of the project consists of research gathering – both from authoritative literature and interviews of Key Informants. The second phase will involve trying to arrive at a preliminary strategy surrounding the use of Big Data in the Department for presentation to senior leadership. To assist with the information gathered, the contractors require an understanding of the current IT structures within the Department. To that end, we are requesting that you or a delegate(s) from your section meet with two of our contractors, Dennis Hogarth and Jacob Sigler as well as a representative from the Research and Statistics Division to discuss issues surrounding Big Data from the IT perspective. We are interested in hearing about the existing IT systems currently used by JUS as well as your perspective on possibilities/challenges on developing and using Big Data. The success of this project rests with being able to speak with representatives from multiple sectors about the practical implications of big data (e.g., IT, Business Analytics Centre, E-discovery – Litigation, Centre for Information and Privacy Law).

The contractors will be in Ottawa next week and we would like to schedule a one hour discussion on June 2nd or 3rd. Please have your office contact Charlotte Fraser, Principal Researcher at (613) 948-3015 or charlotte.fraser@justice.gc.ca regarding your or your delegate's availability.

Attached you will find further background and questions that we would like to address during the discussion.

Thank you,

Suzanne

Suzanne Fortin
Director/Directrice
Priorities and Planning Division/ Division des priorités et de la planification
Policy Sector/Secteur des politiques, Justice Canada
284 rue Wellington Street, EMB/ECE-5287
Ottawa, Ontario K1A 0H8
613-948-3494
suzanne.fortin@justice.gc.ca

Big Data at Justice and Privacy Implications

The Research and Statistics Division, Policy Sector, Department of Justice Canada has contracted with E.S. Tunis and Associates Inc. (http://estaconsulting.org/) to conduct research on big data and privacy. The following two individuals within this firm are leading the project:

Ted Tunis, President

Ted Tunis is President of ESTA and has over twenty-five years of experience as a management consultant, working both within Canada and internationally. He has managed projects in more than 100 international, federal and provincial departments and agencies. Ted has worked at all levels of the workplace where his role has ranged from providing strategic advice to Cabinet Ministers and Deputy Ministers in the Government of Canada and senior UN leaders, to assisting program managers, to engaging office and shop floor employees.

Ted specializes in developing strategic direction and the subsequent management of change, drawing on skills in: organization review, design and development; strategic planning; change management; program design; human resource management; and training design and delivery.

Dennis Hogarth, Senior Associate

Dennis Hogarth has had over 38 years of experience working with KPMG, serving in both client service and internal management roles. For 17 years, Dennis served over 100 KPMG member firms as his clients in a large, multinational corporate environment. His roles in that capacity included the design, implementation and management of organization-wide information, Communications and Technology systems. His work includes considerable experience in the formation, evaluation and restructuring of ICT functions, as well as the recruitment of appropriate individuals to form high performance teams that are effective at assuming responsibility for managing ongoing ICT operations.

Project Overview

Research Phase, consisting of 2 components

- Possible uses of big data by JUS
- Data privacy issues associated with the Department's acquisition and use of Big Data

Strategy Development Phase

Research Phase

Through a review of literature and key informant interviews the contractors will identify the critical issues, and establish a general overview of the key players, the environment, and the current state of the industry in general.

IT Questions to address include:

What applications are in place today that might be of use (e.g., document management, timekeeping, e-discovery, predictive analytics, etc.)?

What are the trends in analytics and big data over the next 3-5 years in the judicial area?

Are there any significant ICT trends in the next 3 to 5 years that might provide other useful sources of data (e.g., cloud computing, mobile, cyber security, etc)?

What will be the ICT needs of JUS over the next 3 to 5 years? How will the demand for ICT applications, tools, and systems evolve?

What are other key external developments in the ICT area expected to impact JUS over the next 3-5 years?

Will the government be focusing on consolidating or decentralizing ICT services in the next 3-5 years?

Strategy Development Phase

A 1 to 1.5 day strategy session is being planned for September, 2015. The session will review the evidence collected in the research phase, identify the strategic issues to be discussed, define the purpose of Big Data for justice and propose options in response to the issues identified. The session will include approximately 20 individuals from Justice representing multiple areas of expertise (e.g., IT, research, privacy law, policy, etc).

Tilhoff, Tanya

From: MacLean, Alyson
Sent: 2015-Jul-28 11:16 AM
To: Akerley, Marj; Sampson, Tracey; Topshee, Dugald; Roy, Dominique; Jakob, David
Cc: Fraser, Charlotte; Li, Ting
Subject: For Comment: Draft Big Data

Importance: High

Follow Up Flag: Follow up
Flag Status: Completed

Further to your participation in a meeting on June 3rd with the Research and Statistics Division and E.S. Tunis and Associates (ESTA) (Dennis Hogarth and Jacob Sigler) about possible uses of big data by the Department of Justice and related privacy concerns, attached is the draft report prepared by ESTA for your review and comment.

We would appreciate feedback by August 7th.

If you have any questions, please do not hesitate to contact me.

Big Data and Privacy Implicati...

Alyson MacLean
Acting Director | Directrice par intérim
Research and Statistics Division | Division de la recherche et de la statistique
Department of Justice Canada | Ministère de la Justice Canada
284 Wellington Street, Room 6119 | 284 rue Wellington, pièce 6119
Ottawa, ON K1A 0H8
Telephone | Téléphone 613-957-9601
Facsimile | Télécopieur 613-941-1845
Government of Canada | Gouvernement du Canada

Possible Big Data Uses by the Department of Justice

And Related Privacy Concerns

Draft Version 2

Version Date: July 24th, 2015

Prepared By:

2B-268 FIRST AVENUE OTTAWA, ON CANADA K1S 2G8

T 613 594 3033, F 613 594 8928

info@estaconsulting.org

www.estaconsulting.org

Field Code Changed

Field Code Changed

Table of Contents

Section I: Executive Summary

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends". What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement.

Big Data is reforming many aspects of today's world, and organizations everywhere are finding ways to use it to achieve competitive advantage. The Canadian government will eventually be obliged to adopt Big Data applications in order to remain internationally competitive. An overall strategy, which considers all of the relevant issues, would help the Government of Canada and all of its departments and agencies to harness Big Data while fulfilling its responsibility to protect the public.

The Department of Justice (JUS) asked E.S. Tunis & Associates Inc. (ESTA) to conduct research into applications and uses of Big Data being made in legal and justice systems that might be considered for use by JUS, and what a Big Data strategy might look like for the department. ESTA was also requested to consider the potential data privacy and protection implications associated with the use of Big Data by the Department. The research method included a review of primary, and secondary sources both internal to JUS and external. Following is a summary of the research findings.

Big Data Applications in the Justice System

Considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data in the legal profession. A large part of the information generated by the legal community or used in court proceedings is in electronic form, but much of this is unstructured – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex; they have taken time to develop and refine, but they are now coming rapidly on stream. In fact, the marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantage and cost efficiencies to those who adopt them.

There is widespread and growing use in western countries of intelligent eDiscovery software tools to analyze and refine large files of relevant documents for the production of evidence to be used in trials. New sophisticated analytics programs are emerging in the US to accurately predict case outcomes without the need to go to trial. Other potential uses of Big Data applications for consideration by JUS might include:

- Techniques to enhance and analyze large operational databases such as the JUS Case Management and Timekeeping systems to improve JUS productivity and cost-efficiency and, in future, to manage resources to address emerging trends.

- The use of data analytics to predict environmental trends using internal (e.g. StatsCan) and external (e.g. social media) information to contribute to policy debates.

- The use of automated tools and Big Data sources for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process.

Promising innovation is also taking place in the justice systems of other countries. Singapore launched a countrywide Integrated Electronic Litigation System (iELS) for all litigation to optimize scheduling of court dates, streamline court filings, and provide case management manage high volume litigation. iELS is accessible from anywhere through an internet browser.

The development and adoption of a Big Data strategy by JUS will not be a simple or inexpensive undertaking. It would require careful planning and long-term commitment if the strategy is to be successful. It will not be possible for JUS to stand still in this area. JUS lawyers will find themselves at a competitive disadvantage to other lawyers in courtrooms, and these pressures will inevitably initiate change. The strategic decision to be made is whether JUS will be an early innovator or a "fast follower". Either way, a careful planning and budgeting exercise will need to be undertaken.

JUS already makes use of technology applications that will provide it with a strong base from which to move forward with the deployment of Big Data and predictive analytics systems. Over the long term, it is predicted that the implementation of Big Data applications will provide both quantitative and qualitative benefits to JUS.

DATA PRIVACY CONSIDERATIONS

Almost by definition, the concept of Big Data in government runs contrary to the concepts of personal data privacy, because a Government Big Data repository must ultimately contain a great deal of personal information about its citizens. Even if a data source is carefully screened to ensure that data is appropriately "de-identified", these protections may disappear when the data is combined with other sources for other uses. This raises potential privacy concerns and, depending on the situation, the potential for negative public opinion.

The implementation of Big Data systems by JUS specifically, and the Canadian government generally, will be challenging from a number of different standpoints.

- There is no single law or practice governing data privacy across Canada; different laws govern the privacy of personal information in the Public and Private Sectors and legislation exists at all levels of government. Although many laws are similar in concept, they are not always aligned, leading to a complex matrix of legislation and practices that surround the use of personal information in the private and public sectors in Canada.

- Some privacy laws also have cross-border and extraterritorial reach. Canada is one of the few countries accepted by the European Union (EU) as having adequate data privacy protections for personal data transfers from the EU. While this status speaks to the strength of Canada's privacy laws, it needs to be preserved as it gives Canada an economic advantage over the many other trading nations that do not have the same status.

- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes. These laws may need change or, at a minimum, to be reconciled as to how they apply in practice.

PUBLIC OPINION ABOUT GOVERNMENT BIG DATA AND DATA PRIVACY

Canadian public opinion about government use of Big Data mainly surrounds how their information will be used and protected. Canadians will be concerned with the security and privacy of their information held by government:

- From an IT security standpoint
- From a transparency standpoint (having knowledge of what is being done with their data).
- From a trust in government standpoint.

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become concerned about their personal information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. The implementation of a Big Data repository by

government is likely to require greater government transparency about the way in which government handles personal information in Canada, and also a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

SUMMARY

There is probably no alternative to the future use of expanded Big Data applications and repositories by JUS. It will become an imperative, if the operation of the Department is to remain cost-effective and competitive. Ultimately, the privacy concerns that arise from the use of Big Data by JUS, on its own, are likely manageable. The implications of the increasing and much broader capture and use of Big Data by government in general creates a number of legal, policy and other issues that JUS will inevitably need to help to resolve as it moves forward with Big Data applications.

Section II: Introduction

II-1: The Evolution of Technology and Big Data

The explosion of computing, electronic sensing and digital communications technology in today's society has led to an exponential growth in online data; we create roughly 2.5 quintillion bytes of data a day, so much that an estimated 90% of the data currently in existence were created in the last two years (IBM, 2015). Large, growing subsets of this mountain of data are referred to as Big Data.

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends. In particular, these findings relate to human behavior and interactions. For the most part, these are datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Brown&Ehrenreich, 2015). What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement (Ward & Barker, 2013).

Big Data repositories are a result of the exponential increase in the amount of data being captured, combined with advances made in low cost digital storage media. Almost all transactions are now done online, and most documents and forms are now available in digital form only. Internet-enabled devices that are capable of capturing personal, environmental and geolocational data surround us. This data is being used and combined in increasingly innovative ways that were often not anticipated during the initial collection process. Governments and private organizations alike are beginning to recognize the value of this data, and are investing heavily to harvest it to gain a competitive edge and other strategic advantages.

As data repositories have expanded and evolved, so too have the methods and processes that permit data search and manipulation. The cost of storage has decreased to the point where much data is kept indefinitely, often because it is easier and cheaper to do so than to devote resources to culling it. In the meantime, advances in processing power and the creation of new ways to combine and analyze the data have permitted the combination and parsing of the data for novel uses.

This new flood of information has led to a large number of opportunities across a wide variety of sectors, while at the same time giving rise to some new privacy concerns as more data is gathered in a world where many electronic devices are now internet enabled, and are beginning to monitor and store information on almost everything that we do. Given enough data about an individual, it is possible to create a very detailed profile that removes all prospects of future privacy.

II-2: The Inherent Conflict between Big Data and Data Privacy

The data that exists in a Big Data world must ultimately include a great deal of information about real people. In the past, this information existed in siloes that were, for the most part, physically separated because information was stored in paper files. Even in the early days of electronic data processing, information was stored on devices that were only accessible by individual computers with no connection between them. In today's world of Big Data, these electronic files are capable of being linked both physically and logically together to permit broader information access and greater system functionality.

Clearly, there is growing value in harnessing Big Data. Predictive modelling using Big Data sources will permit doctors to make more accurate medical diagnoses (Dwoskin, 2014). Medical diagnostic programs may soon be capable of using Big Data findings to review a patient's entire medical history, X-Rays and results of medical tests online – from virtually any location in the world – to make a diagnosis. Vendors can use multiple sources of information to predict retail trends and match their supply of goods and resources with anticipated demand. Governments can monitor health and other emerging social trends in their countries to forecast the need for public programs, resource allocations and budgeting.

An individual's privacy has long been considered a fundamental human right. However, the Canadian Charter of Rights and Freedoms, when enacted in 1982, didn't anticipate a world where an individual's personal information could be captured and stored in such minute detail, nor the ways in which it might need to be specifically protected. Sections 7 and 8 of the Charter have often been interpreted to provide these protections, but may not provide the required degree of specificity in a world where the various permutations and combinations of the data make it very difficult to ensure individual anonymity.

One of the first big uses of analytics applied to Big Data sources in government has been by the intelligence community, which developed programs such as Carnivore[1] to monitor and analyze large amounts of electronic communications in order to detect subversive activities. Predictive analytics are also being used to forecast crime levels based on regional and local demographics. This information is also being used, primarily in the US, in predicting an offender's likelihood of reoffending as a basis for sentencing decisions. Big Data history is already being used to predict future population trends. As more data is captured about the everyday activities of individuals, it will not only be possible to make predictions about their health and welfare as a basis for improvement, but also whether they may be more susceptible to committing criminal acts before they commit them. The Big Brother world of George Orwell's "1984" might have arrived.

[1] Carnivore was a system implemented in the US in 1997 by the Federal Bureau of Investigations to monitor email and electronic communications sent over the Internet

While there are ways to disguise personal information in large data sources, the current focus of investment and research is mainly on ways to harvest Big Data, rather than on how to protect it, along with an individual's data privacy. An appropriate balance will need to be established if Big Data and privacy are to co-exist peacefully in Canadian society.

II-3: Why Big Data?

Regardless of the challenges, Big Data offers considerable opportunities to the Department of Justice (JUS). In order to take advantage of the opportunities, and to minimize the negative effects of Big Data, JUS needs to develop a clear picture of the current state and likely near-term evolution of the technology. To this end, JUS commissioned ESTA Consulting to conduct research into:

- The impact of Big Data in the context of current privacy laws in Canada;
- Ways in which JUS could adopt Big Data for its own needs;
- The implications/opportunities of Big Data including the possible role for JUS, and whether a "Big Data Strategy" would help; also more generally for the Government of Canada.

Implementing a Big Data strategy is not a simple task, especially for organizations the size of JUS or other federal public departments. All organizations now use Information Technology (IT) to a greater or lesser extent, but it is important for organizations to understand their current use of technology as a prerequisite for planning how they might move forward. The purpose of this report is therefore threefold:

1) To broadly review the current applications of IT by the Department of Justice in order to help it assess its position vis-à-vis other legal organizations with respect to the implementation of the advanced technologies and techniques employed by others to harness the power of Big Data in the public and private legal sectors;
2) To identify some of the privacy issues from a legal or regulatory standpoint that might stem from the availability and use of Big Data by JUS and the federal government, both currently and in the foreseeable future;
3) To consider the implementation of Big Data by the Government of Canada and some of the broader issues that could arise from this use, including public opinion and reaction.

Section III: Methodology

III-1: Project Scope

JUS requested the following scope in the form of questions to guide the direction of the research:

SECTION 1: HOW THE DEPARTMENT OF JUSTICE CAN MAXIMIZE THE USE OF BIG DATA?

Q1. Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

Q2. How can the Department of Justice adopt Big Data for its own needs?

Q3. Are there promising practices in other countries and departments worth emulating? Where and what are they?

SECTION 2: PRIVACY

Q4: What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

Q5. What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms, incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms.

Q6. What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

III-2: Research

The following research activities were undertaken:

INITIAL RESEARCH

Initial research was performed to assist with the scope of the research, and in planning. This involved an initial review of material available online, and meetings and discussions with JUS research staff members to clarify roles and responsibilities. Initial research also included a literature review to identify existing and near-future uses of Big Data in the legal sector, both

private and governmental. The literature review also helped to better define the scope of the research.

PRIMARY RESEARCH

Primary research consisted of interviews with Key Informants, both internal and external, to identify issues and help establish an accurate overview of the industry, as well as to assist in determining the criteria to be used in analyzing the results and developing the conclusions. Selected Key Informants' views were solicited to help shape some research, highlight background issues and subject matter, and provide some assistance with key observations and conclusions. Their comments, where relevant and notable, were included verbatim in the report. (See Appendix XI-7~~XI-7~~ for the complete list of Key Informants interviewed).

Formatted: Hidden

Key Informants were also asked for their views and observations on the subject of potential Big Data uses in JUS and the Federal Government, and on the associated data privacy issues and public perception, in order to identify the issues and to establish a general overview of the environment and the potential issues and concerns. A Key Informants plan and guide was assembled to direct discussions with the informants, but questions were modified for each interview to match the particular area of expertise of the Key Informants. The focus of the questions was on the direction of Big Data development in the legal marketplace, and possible data privacy implications associated with the use of Big Data, both by JUS, and more broadly by the Canadian Government.

In order to gain an understanding of Canada's use of Big Data in relation to the rest of the world, research was also conducted into the uptake of the identified technologies in various jurisdictions. An overview of other ways foreign governments use Big Data was also established.

SECONDARY RESEARCH

A broad background and view of the environment, drivers, issues, and industry players was developed from the initial and primary research. Published reports, research papers, websites, Internet sources on the topics, together with media reports, were then examined. Further research was then conducted into each of the identified Big Data uses in order to understand their capabilities, limitations and methods of use, and to identify the most common product options in use. Secondary research focused on areas of legal administration related to the business of JUS.

All research was conducted with a view to produce an initial identification of emerging issues and risks in the use of Big Data, primarily by JUS, but also more generally by government agencies.

Drafts and the final reports were reviewed with JUS staff to ensure accuracy and to verify scope coverage.

Section IV: The Emerging Uses of IT In the Field of Law

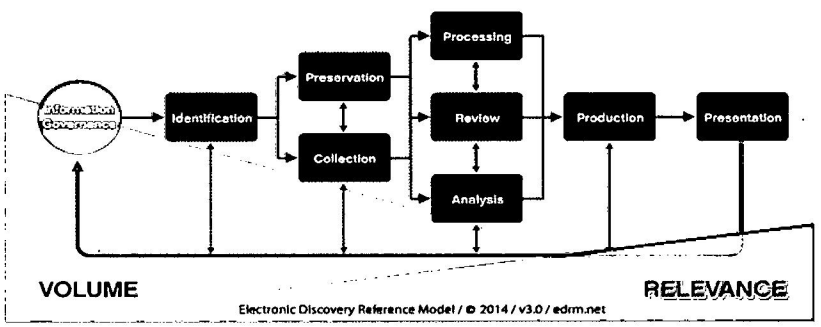
While technology applications, such as practice management systems (e.g. for time-keeping and financial management) have been used for some time in legal service organizations, the broader use of technology tools has been a relatively recent development. This has likely been driven in part by the explosion in the amount of unstructured (i.e. text-based) information in electronic form, and partly by innovations in the technology world to improve the ability to search, correlate and interpret this unstructured information in meaningful ways to gather and interpret evidence used in legal cases.

This section discusses some of the rapidly evolving uses of technology in the legal profession, including enhancements attributable to the emergence of Big Data; the sophisticated tools used to analyze large stores of data in the areas of eDiscovery and evidence gathering; legal research; the prediction of trial risk and outcomes and in the use of advanced data analytics for practice management and to achieve productivity improvements.

IV-1: eDiscovery Methodology and Tools

Electronic discovery (eDiscovery) tools include software designed and used to identify, preserve, collect, process, review, analyze and ultimately to produce information in electronic form to support the legal discovery process as legal cases are being conducted. eDiscovery software capabilities include the ability to identify, preserve, collect, process, review and produce information for use by counsel. These capabilities are generally conducted in a sequential order prescribed in the Electronic Discovery Reference Model (EDRM, 2015), a framework that has been established and is broadly accepted by eDiscovery practitioners.

Electronic Discovery Reference Model



During the initial phases of the eDiscovery process, the data collection and early assessment capabilities of eDiscovery software is used to refine the data so that an initial evaluation can be made regarding the information quality, the location of information that is available for use in a case, and what additional resources might be required for its effective evaluation. A risk assessment is generally performed during this stage to determine whether any restrictions might govern the use of the data, such as policies or data protection laws.

Subsequent phases of the eDiscovery process generally include technology assisted review tools that employ analytics-based machine learning technology. These use statistical techniques to "train" the software to review the electronic files, thus reducing the required amount of manual review to improve overall cost-effectiveness of the review process.

eDiscovery tools have evolved considerably since they were first introduced to the legal marketplace. In its May 2015 "Magic Quadrant for eDiscovery Software" study, Gartner Group (Gartner Group, 2014) studied 18 of the top organizations providing eDiscovery solutions and services to the marketplace today. They positioned the 7 organizations described in Appendix XI-1~~XI-1~~ as the current industry leaders.

Formatted: Hidden

Key Informant Kelli Brooks, who heads up KPMG's Evidence and Discovery Management Group in the US, noted that the kCura Relativity platform is the most commonly used tool, but indicated that the following eDiscovery platforms had potential for creating significant developments in the eDiscovery industry:

- Equivio is a relatively new Israeli text analysis start-up company that was bought by Microsoft in 2015. Industry speculation is that Microsoft plans to integrate the Equivio machine learning technology into Office 365 in future.
- Brainspace is a revolutionary new tool that can be used to reveal complex relationships between documents for review.

PREDICTED CHANGES IN THE EDISCOVERY MARKETPLACE

Transparency Market Research, a U.S.-based provider of syndicated research, customized research, and consulting services estimated that the Global eDiscovery market was valued at USD 5.56 billion in 2013. Government and regulatory agencies were the largest end-user segment in 2013, accounting for about 51% revenue share of the global eDiscovery market. They expected the market to grow at a cumulative annual rate of 15.5% from 2014 to 2020 as eDiscovery solutions find widespread applications in government and regulatory agencies, small, mid and large-sized enterprises and law firms. (Transparency Market Research, 2014)

The eDiscovery marketplace will also change as electronic evidence expands from the current analysis of email, documents and voice mail to include social media and mobile data. Increases

in data transfers between inter-connected business systems will require growth in the ability to analyze structured data. A combination of human skill and sophisticated software tools such as predictive coding and structured data analytics will be required to analyze these more complex evidence streams. A number of large players in the IT world are investing heavily in both eDiscovery software and tools for predictive analytics in the legal marketplace. These include:

HP Autonomy – The Hewlett-Packard purchase of the Autonomy search engine in 2011 for $10.3 billion set the stage for their entry into the eDiscovery marketplace, and they have continued to invest heavily since in new functionality (e.g. a cloud-based offering) to expand in the legal marketplace;

ROSS – a result of collaboration between the University of Toronto and IBM, using IBM's Watson Artificial Intelligence engine for legal research (Krasnyansky, 2015);

Microsoft's purchase of the rapidly growing Equivio in January, 2015 for a rumoured $200 million gave them access to "a provider of machine learning technologies for eDiscovery and information governance. We are making this acquisition to help our customers tackle the legal and compliance challenges inherent in managing large quantities of email and documents." (Microsoft acquires Equivio, 2015).

Key informant Dera Nevin advised that two important issues must be addressed before an organization can move forward with plans to capitalize on the use of Big Data for eDiscovery or more sophisticated applications (e.g. Artificial Intelligence (AI) and Predictive Analytics):

1) An appropriate information governance structure must be in place so that the organization has knowledge of what electronic information they have and where it is stored. In the past, legal organizations have been overly reliant on the use of paper documents, and a significant cultural change is required to overcome this issue.
2) Organizations need to standardize on a limited set of eDiscovery tools to permit legal counsel to become experienced with their use. Lawyers won't become experts in programming, but they will need to become adept in future at using sophisticated tools to search for and manipulate data.
   Although software standardization is a desired goal, Ms. Nevin observed that large legal organizations like the Department of Justice are also exposed to a wide variety of legal scenarios, and since there are specific strengths and weaknesses of the various tools on the market, a single eDiscovery solution might not be suitable for every case. The need to differentiate between structured and unstructured data may also require different tools.

STATUS OF THE USE OF EDISCOVERY TOOLS IN JUS

JUS IT representatives indicated that the Ringtail tool, from FTI Technology, is used across the Government for evidence management. In addition to JUS, the RCMP, PCO, PPSC, Election Canada, and the Treasury Board apparently use Ringtail. However, at least one of the key informants we spoke to expressed the view that JUS use of the tools was not as extensive as it might be, and that JUS might be lagging the private sector in this area.

Jean-Sébastien Rochon and Julie Roy of the National eDiscovery and Litigation Support Services group indicated that about 14-16 paralegal positions are devoted to the support of Ringtail and eDiscovery tools. Ringtail is only used for files involving more than 5000 documents as it is not cost-effective on smaller cases.

A problem highlighted with the current implementation of Ringtail is that documents from the 1700 cases stored in the system are held in "silos", so documents used for evidence in one case aren't available for use in others, although they might be useful. Going forward, Rochon and Roy hope to restructure the Ringtail database so that over 25 million pages of documents could be searched across the system and made available if they are relevant to other cases, and aren't subject to legal privilege.

Another problem they identified was that legal units assigned to other government departments sometimes use other eDiscovery tools not recommended by JUS. These create files that aren't compatible with JUS and therefore can't be shared.

USE OF EDISCOVERY IN OTHER JURISDICTIONS

A review of other jurisdictions finds mixed approaches to the application of eDiscovery. Table 1 (below) shows an overview of the use of eDiscovery and governing laws in various countries:

Table 1: EDiscovery Around the World

Country	eDiscovery Legislation	eDiscovery Use
Canada	• Sedona Canada Principles Addressing Electronic Discovery (1st ed 2008, 2nd ed 2015) (Federal, compatible with all provinces and territories except Quebec, based on US) • Ontario, Nova Scotia, Manitoba, Saskatchewan, Alberta and BC all have guidelines for eDiscovery based on the Sedona principles • Quebec, as a civil law province, has different rules	• Widespread • Following the American example

United States	• Legislation in effect since 2006 (meet and confer), updated 2007, 2015 (pending)	• Widespread • Pioneering and exporting eDiscovery to the world
United Kingdom	• Legislation in effect since 2009 (meet and confer), updated 2013 • Very specific eDiscovery guidelines and requirements • Litigation budget is required early in the process	• Widespread • Some jurisdictions require all cases to use eDiscovery, some allow the judge to make the decision on a case by case basis
Australia	• Legislation in effect since 2009 (meet and confer) • Update in judge training program includes managing eDiscovery and electronic case management	• A court can order all discovery for a case be done electronically. • Most courts have implemented individual guidelines specifically for eDiscovery
New Zealand	• Legislation in effect since 2012 (meet and confer) • All discovery is now electronic, unless the court decides otherwise.	• Waited a long time to make rules, and had a chance to see what other commonwealth countries did • EDiscovery is now ubiquitous
Japan	• No laws governing eDiscovery for domestic litigation.	• Not very common in non-governmental cases • An expectation of data production exists for government investigations • Slowly gaining popularity, mostly driven by international litigation, particularly with US law firms and vendors
Korea	• No specific eDiscovery laws • Very strict privacy laws, including a requirement that all corporate and personal data be hosted domestically	• Virtually non-existent
Singapore	• Integrated Electronic Litigation System (iELS) implemented in 2013	• All cases use eDiscovery through the iELS

IV-2: Timekeeping, Document and Case Management

JUS has used its proprietary iCase tool for a number of years to store documents used in litigation. iCase is also used for time and case management. The JUS IT group indicated during our interview that a major goal is to align JUS systems to the extent possible with prescribed federal government standards. GC Docs has been adopted as the standard for record keeping and document management, with Microsoft SharePoint 2013 as the front-end interface and

system portal. JUS will be converting, but implementation is only in the early stages, with migration of existing content occurring in the later stages.

IV-3: Legal Research

Internal and external information sources are used for legal research. *Justipedia* is the central legal knowledge management repository for the Department of Justice. It contains legal opinions, pleadings and facta, agreements and other precedents and tools. It is also used to access legal practice tools and models, legal training materials, a directory of expertise and other materials. Content is organized by practice area and content type and is searchable.

Access to external published research and data sources for evidence gathering is available through a third party legal research tool called LexisNexis Quicklaw, which gives lawyers access to a comprehensive collection of primary and secondary legal research materials, court decisions, legislation, legal commentaries, and current and archived news

IV-4: Evidence Gathering

While iCase has previously served as the government standard for assembling case documentation for evidence gathering, it is to be replaced by Microsoft's CRM Dynamic. The legal service unit of the Canadian Food Inspection Agency is already using CRM Dynamic successfully for this purpose.

JUS IT representatives indicated that there is a need to identify a faster content search engine for use by the Department; they are investigating the adoption of the Fast Search capability incorporated into Microsoft SharePoint 2013 as a possible solution. This would permit enterprise-wide indexing and search of JUS content and documents in any other repositories to which they have been granted access. Fast Search could potentially be used to create a cross-government Big Data search capability extending beyond JUS itself. During content processing, information can be written to a link database for subsequent use by an analytical capability in the software to calculate link popularity statistics and to perform relevance weighting of documents found. This could make relevant content more quickly available to JUS lawyers, improving their ability to assemble evidence to support their cases.

IV-5: Business Analytics

JUS has a Business Analytics group that uses SAS (Statistical Analysis System), a software suite developed by the SAS Institute that is used for advanced analytics, business intelligence, data management and predictive analytics. SAS can be used to retrieve and modify data from a variety of sources for the purpose of performing statistical analysis.

SAS Analytics is the main tool that is used to analyze data inputs from the various resource management tools in use in JUS, including IFMS, Peoplesoft, iCase, and other sources. Toundjer, Erman, the Director Business Management Strategic Planning and Business Management, believes that while JUS systems that provide operational information and statistics are functional, different systems produce different results. The current focus is therefore on fixing the data before moving forward with plans to enhance the systems to generate more meaningful data. The existing iCase timekeeping system is used for performance measurement.

Problems with the current environment that need resolutions as a precursor to implementing a Big Data approach in the business analytics area include:

1) There are some significant gaps in the current information:
   - An intake system is required to measure the demand for services;
   - The litigation system is not treated as a process, and therefore it is difficult to determine who is adding value;
   - There are 160,000 files on iCase, but a number of these are duplicate entries, or are initiatives that do not represent actual legal cases.

2) Non-chargeable hours aren't tracked, such as the provision of advisory services to clients, so the analysis is incomplete.

3) It is difficult to develop Key Performance indicators because of differences between reports and inconsistencies in the data that is reported.

4) Reliable data isn't available from the private sector for comparison regarding efficiency and performance of the department;

5) There is some internal resistance to providing the necessary data.

IV-6: Big Data Analysis

The JUS IT Department is investigating the use of various tools to perform Big Data analysis – such as HP's Autonomy which allows analysis of large scale unstructured Big Data repositories, and ROSS, an experimental artificial intelligence system built on IBM's "Watson" artificial intelligence platform developed by researchers at the University of Toronto. Although both systems hold promise for the future, they are still at very early stages in their development; any practical implementation of the tool is unlikely to occur for some time to come.

Section V: General and Future Trends

V-1: Future Trends: Possible Big Data Applications in Justice

Research has traditionally involved two fundamental steps - developing an initial hypothesis and finding proof that confirms or refutes the hypothesis. While this approach remains an appropriate research methodology, a new approach has emerged in the world of Big Data. Using artificial intelligence, massive stores of data can be searched for areas of correlation without using an underlying hypothesis previously identified by researchers. As an example, researchers used Google's intelligent search engines to identify a correlation between queries in its Google Trends web site and seasonal outbreaks of influenza in various countries (Google, n.d.).

Similar correlations are beginning to be discovered in the legal and judicial environments. For example, the correlation among outcomes of legal cases, judgments and appeals are beginning to provide the capability to predict the outcome of future cases. Also the correlation between massive stores of case evidence searched in electronic form by eDiscovery tools will provide key findings and evidence trends for use by legal counsel in trials.

Kevin Quinn, a former Assistant Professor of Government at Harvard, ran a contest comparing his statistical model to the qualitative judgments of 87 law professors to see which could best predict the outcome of all the US Supreme Court cases in a year. The law professors knew the jurisprudence and what each of the justices had decided in previous cases. They also knew the case law and all the arguments. Quinn and his collaborator, Andrew Martin collected six crude variables assembled from previous cases and analyzed the outcomes, which exceeded the lawyers' predictions. They concluded that whenever sufficient information can be quantified, modern statistical methods will outperform an individual or small group of people. (Shaw, 2014)

V-2: Predictive Analytics and Early Case Assessment

Lawyers make many strategic decisions and predictions during any stage of a trial based on their assessment of the outcome. Lawyers may also decide before taking a case to trial whether to negotiate a settlement offer. The ability to accurately predict the outcome of a case has practical consequences because litigation is risky, time consuming, and expensive. Errors in judgment can be costly in terms of time and resources, and also place a significant burden on the judicial system.

Many large legal firms are adopting the use of early case assessment tools and methodologies to estimate the risk of prosecuting or defending a legal case based on the financial costs and

resources required. Electronic legal discovery is also becoming increasingly costly. Organizations that spend significant resources on a case may eliminate the cost benefit of going to trial. Some organizations are also using the volume of information that can be produced to make cases more difficult and costly for the other side of the case to prosecute or defend.

Some existing software tools that can assist in and help facilitate the process of early case assessment include eDiscovery tools such as Exterro and Open Text eDiscovery. A US-based software company has also developed an application called "Picture It Settled", another example of a software tool used for early case assessment. This tool apparently uses neural networks, probability theory and behavioural patterns to predict the actions of opponents in a case, which can help to streamline negotiations. The software also estimates when parties are likely to settle and for what amount, with high accuracy. This doesn't replace legal judgement, but helps to understand alternatives and guide decisions by quickly modeling anticipated reactions.

Effective early case assessment requires a combination of professional expertise and software. Different resources in an organization typically use the software to assist in analyzing both structured and unstructured information[2] stored in electronic form. Depending on the sophistication of a case, lawyers may be assisted by IT professionals, forensic teams, and independent consultants. The tools used and the results of an early case assessment review can vary. Early case assessment is not a "one size fits all", but rather a process that needs to be managed and customized for each case.

The use of Big Data for case settlement and alternative resolution is expected to be one of the most significant future uses of Big Data in the judicial system. Information produced by the Data Analytics group in JUS indicated that the majority of cases processed by JUS are relatively small; in fact large cases are the outliers in statistical terms. While some cases processed by JUS must be taken to trial, many small cases may go to trial where the outcome can be predicted in advance. Significant savings in settling those cases without having to go to trial might result.

While JUS might be obliged to take a case to trial on principle, regardless of the possible outcome, predictive analytics may offer the opportunity to avoid trial in many situations.

[2] Structured data is organized in a highly mechanized and manageable fashion which can be easily processed by a computer, such as stored in Excel spreadsheets; by comparison, unstructured data, such as text found in e-mails and text reports is raw and unorganized. Searching through unstructured data can be expensive and difficult.

V-3: Current Limitations with Prediction Models

There are currently limitations to the predictive analytics approach to case outcome prediction and/or settlement. Predicting the outcome of new legal cases is still an imperfect science because of limitations of the current information is available for inclusion. e.g.:

- Cases may be settled without going to trial and aren't available for inclusion in the database, making the data incomplete;
- Courts may not have decided enough similar cases to permit the statistical prediction of case outcomes or feature weights that are needed to resolve the problem of small or biased samples;
- Algorithms that rely solely on assigning quantitative feature weights can be problematic because they are not sensitive to the particular context of a problem;
- The statistical algorithms used in the prediction models require sufficiently large data sets and, the more difficult the task, the more cases are needed to achieve accuracy;
- Text cases need to be represented in an appropriate form to enable machine learning; this is currently a largely manual process.

These difficulties are likely to be overcome with time and, given an appropriate database of cases, statistical or symbolic machine learning[3] techniques will be used effectively to determine general rules for classifying new cases and predicting their outcomes.

One major impediment to predictive analytics faced by JUS and the Canadian legal profession is the expense of building a complete and accurate Big Data store of cases and precedents. The information must also be kept current for new legal decisions and appeal results. It is unlikely that such a project could be funded in the near future without the backing of a consortium of law firms, or a third party organization such as LexisNexis who might make the information available by subscription. However, a detailed cost-benefit analysis would need to be performed before embarking on such a large project.

As a comparison, new regulations governing the accounting profession in 1999 forced the large accounting firms in the US and internationally to commission the development of a database containing information of all public and private companies, for use in determining possible conflicts of interest impairing auditor independence. Collectively, the firms engaged Sentinel, an organization supporting brokerage firms, to augment and modify their existing database of public and private organizations, and associated systems tools to accomplish this objective.

[3] Symbolic Machine Learning is another term used for predictive analytics or modeling where patterns of data are identified using human readable terms and symbols as opposed to numbers.

V-4: Data Management and Analytics

Controlling the information that is captured in large datasets can be problematic and subject to legal or ethical restrictions including:

- Documents used as evidence that contain personal information;
- Third party sources of information, such as articles or agreements that may be subject to copyright laws preventing open disclosure or dissemination;
- Confidentiality agreements where open disclosure could cause harm to a third party;
- Content compliance with government policies and practices.

Content management and curation of a JUS Big Data site will be an onerous task. Data will need to be kept current, as well as in compliance with laws and policies. Fortunately, software tools are being developed to assist with this process in the form of Data Management Solutions for Analytics (DMSAs). Gartner Group describes a DMSA as "a complete software system that supports and manages data in one or many disparate file management systems (most commonly a database or multiple databases) that can perform relational processing (even if the data is not stored in a relational structure) and support access and data availability from independent analytic tools and interfaces" (Gartner Inc., 2015). Organizations offering these tools include traditional IT firms, such as Teradata, Oracle, IBM, Microsoft, SAP and HP. However, new organizations, such as Cloudera, MapR, Actian and Pivotal are competing with the leaders.

Toundjer Erman, indicated that his objective was the "integration of information from all JUS systems that generate Enterprise Resource Management information in order to get a holistic view of all JUS operations." In parallel, there is a need to consider what the new operational landscape should look like, and then to generate new ideas by "looking through different lenses" and gaining new insights. This would include taking into consideration what other governments and public sector organizations are doing to use Big Data and technology to improve legal service processes and efficiency.

Ultimately, existing JUS data analytics information could be combined with other data for use in predicting how the legal environment will change. For example, will new legislation trigger more litigation, and what resources will need to be recruited or developed in JUS over a period of 3-5 years to respond to those predicted needs. Predictive Data Analytics can contribute to this analysis, but will require redevelopment of the current data architecture in the administration and resource planning areas to be more process driven.

V-5: Policy Development

Big Data offers an opportunity to contribute to government policy debates. Tools such as "Social Harvest" can extract data from Twitter, Facebook, and other social media platforms and log this information to a variety of data stores. Statistics Canada and many other government agencies possess a wide range of data concerning the behaviour of Canadians as a direct result of citizen interactions with government online services. However, a government department that uses social media to try to identify and better understand the needs of Canadians might also be accused of spying on its citizens in order to supress potential resistance.

The use of Big Data for policy development raises new moral and ethical issues for policy makers. Using predictive analytics and probability theory to predict what the general population might do in the future, as opposed to what they have done in the past could contribute to the policy debate. However, results based on findings from a relatively small group of people might still contain errors. A risk is that Big Data predictions about individuals might punish people for their propensities, not their actions, thus potentially denying basic human rights. Predictive analytics used by police in the US has led to a reduction in certain crimes, but resulted in the targeting certain socio-economic or cultural groups. (Joh, 2014)

Section VI: Other Government Big Data Sources and Uses

Big Data offers a wealth of opportunities for other government agencies. Table 2 (below) shows a few of the areas in which Big Data is being exploited by other governments around the world.

Table 2: International Governmental Uses of Big Data

	US	UK	Australia	New Zealand	Singapore	Israel	Other
Predictive Policing	✓	✓	✓	✓	✓	✓	Spain
Informed Sentencing	✓	✓		✓			
Bail/Parole	✓	✓	✓			✓	
Fraud Detection	✓	✓			✓	✓	Canada
Health Care	✓	✓			✓	✓	Taiwan
Education	✓	✓		✓		✓	Korea, Canada
Public Works	✓				✓		Ireland, Philippines
Transportation	✓		✓	✓	✓		Sweden, Ireland
Infrastructure				✓	✓		
Economic Policy			✓				Japan, Germany, Canada
Environment				✓			Netherlands, Canada
Public Relations					✓		Japan, Hong Kong, China
Information Sharing	✓	✓			✓		Spain, Ireland, Japan
Government Resource Allocation			✓	✓	✓		Philippines, Germany

VI-1: Sentencing and Parole

Big Data can have a big impact on Correctional Services and on the Criminal Justice system in general by informing sentencing and parole decisions in a variety of ways.

Data-centered, evidence based strategies can be used to divert as many people as possible toward alternative programs, either within or outside of prisons, possibly reducing prison crowding and lowering the likelihood of re-offense. The Attorney General of the United States says "[d]ata can [...] help design paths for federal inmates to lower these risk assessments, and earn their way towards a reduced sentence, based on participation in programs that research shows can dramatically improve the odds of successful re-entry. Such evidence-based strategies show promise in allowing us to more effectively reduce recidivism" (Leopold, 2014). Similar risk assessments can be used to inform bail and parole decisions.

These types of strategies are being used effectively in a variety of jurisdictions:

The State of Florida and the province of Quebec both use statistical programs to profile juvenile offenders and assign them to risk-specific rehabilitation programs. These programs have shown significant success in reducing recidivism (Perry, McInnis, Price, Smith, & Hollywood, 2013);

The US states of Pennsylvania and Tennessee and the Australian state of New South Wales require statistical analysis to be used in all sentencing decisions;

The cities of Baltimore, Philadelphia and Washington, DC, all use algorithms to predict the likelihood of re-offence by parolees, and plan parolee supervision accordingly.

Big Data can also be used at a higher level to inform overall sentencing guidelines; the US Sentencing Commission is currently studying the use of data-driven analysis to issue general (not individual) policy recommendations. These could include changes in recommended sentence length where historical data shows current measures to be ineffective.

VI-2: Policing and Security

Law enforcement agencies have a history of using profiling and data mining to identify potential threats and predict criminal activity: Big Data offers a variety of tools to augment this capacity.

DEPLOYMENT

Predictive analytics are being used in over sixty major cities across the United States to help law enforcement agencies predict areas of probable criminal activity, and to assign patrols accordingly. These programs take into account times and locations of previous crimes, incident records, weather patterns, and historical and sociological information to create maps of "hot

spots". Cities using these maps have reported decreases of between 10% and 40% in criminal activity as a result. Los Angeles also tweets daily "hot spots" to citizens, to increase vigilance.

CRIME PREDICTION

Predictive analytics can also be applied more narrowly, to identify individuals at high risk of committing crimes. Chicago has a program in effect that uses a "heat list", created by a complex algorithm using data from a wide variety of sources. Officers or letters are sent to the homes of people on this list, to offer social services such as job training, or tailored warnings of increased penalties for certain crimes for people with particular prior convictions. The program has yielded positive results and is considered a success.

The U.S. Department of Homeland Security (DHS) and the Israel Security Agency (ISA) both have programs under development to detect terrorist attacks before they happen. DHS uses a Future Attribute Screening Technology to screen people for behavioural attributes associated with violent acts. Their Predictive Screening Project defines observable behaviours that precede a suicide bombing attack, and has shown promise in the testing phase. The ISA is investing in technology to convert unstructured data such as video and audio into a form that can be analyzed and used to produce real time alerts.

CRIME DETECTION

A third area of use for Big Data in policing is detecting crimes in near real time. This is being applied mainly to various forms of fraud, such as Medicare, securities, and bank fraud in the U.S. It is also being used in the UK to detect the misuse of prescriptions, and foreign bribery.

VI-3: Full Litigation Services

In 2013, Singapore launched a country-wide Integrated Electronic Litigation System for all litigation. iELS is accessible from anywhere through an internet browser, and has the following key functionalities (Braddell Brothers, 2015):

- Streamlining and re-engineering of high volume litigation processes;
- Information-based filing - Data capture (e.g. via XML and electronic forms) instead of only paper capture (e.g. document scanning), enabling the flexible re-employment of information as and when required;
- Active case management - Courts can pro-actively track and manage pending matters
- Litigation process management - Alerts and triggers designated to ensure that litigants do not miss critical deadlines;
- Electronic case file for lawyers - Lawyers have access to all relevant documents at any time and any place with an Internet connection, for the duration of each case;

- Integrated due diligence checks - Due diligence checks integrated with the electronic filing process, doing away with the need for subsequent back-room reconciliation;
- Court calendaring - Optimal assignation of court hearing days to be achieved with the syndication of date/scheduling information captured via information-based filing.

VI-4: Transportation

Intra and inter-city transportation systems (including both infrastructure and services) produce a vast amount of data from sources such as road sensors, bus GPSs, and ticketing systems that can be analyzed and used to increase the efficiency of services and allocate government resources. Some examples of foreign governments using these data to great advantage include:

- Swedish National Road Administration uses IBM systems to predict, control and optimize road traffic to improve air quality and reduce congestion. This resulted in peak-time road traffic congestion being dramatically reduced, air pollutants cut by up to 12 percent, and public transport usage increase significantly;
- The city of LA uses demand-responsive pricing for parking. Prices are based on data from parking sensors, surveys, weather forecasts, information about holidays, local business activities, etc.;
- The city of Dublin provides live road sensor and city bus GPS data to citizens, who can use it to plan their routes;
- Similarly, New Zealand uses predictive analytics to provide motorists with real-time information on traffic patterns via Variable Message Signs, in operation on highways across the country. These signs also display messages about accidents and road closures and conditions.

VI-5: Health Care

A large variety of health related data exists (patient records, genome information, successful/unsuccessful trials, hospital records etc.). By combining this data for analysis, variants of a disease can be identified, as well as subsets of patients who would benefit from different treatment plans. Following up with these groups could lead to better outcomes for the patients, and greatly advance the research, although this can be difficult if information is anonymized or de-identified. (President's Council of Advisors on Science and Technology, 2014)

Some examples of Big Data currently being used in the health care field include:

- New Jersey uses medical billing data to map out hot spots where there are the most complex and costly healthcare cases, as part of a program to lower healthcare costs
- The UK Food Standards Agency uses Twitter data to predict outbreaks in real time (often weeks before other methods);

- In Singapore, hospitals are using predictive analytics to predict relapses;
- Taipei Medical University analyzes and monitors performance across all hospitals.

VI-6: Economics

Reliable information about the current state of the economy is extremely important in making monetary policy decisions. Big Data can provide this information by predicting a wide variety of econometrics. For example, there are a variety of leading and lagging indicators of overall unemployment in a jurisdiction, such as automobile downgrades and decreased grocery spending (leading), and increased foreclosures and vacation cancellations (lagging). This sort of analysis can be used for early warning, real time awareness, and real time feedback for public policies and programs. (Letouze, 2012)

The Bank of Canada has suggested using existing monthly indicators in combination with big data to predict GDP growth before official quarterly National Accounts data are released providing more timely and accurate metrics to inform monetary policy decisions. (Armah, 2013)

VI-7: Education

With the advent and increasing popularity of online learning, there are new sets of data available about how and what students learn, including responses to various new techniques and modes of delivery. Research into these data could yield great benefits to the field of education, including identifying what skills taught at which points in childhood, leading to better adult performance in certain tasks. Learning management systems (for use in actual classrooms) are also becoming more popular, and are adding to the available data. (President's Council of Advisors on Science and Technology, 2014)

Student data can also be used to identify and respond to student having educational difficulty. In 2012, Ontario's Ministry of Education identified 14,000 students across the province who had left high school with three or less credits needed to graduate. One year later, after a campaign to get them to go to summer school or take extra credit courses, 8000 of them had graduated. (Solomon, 2013)

Section VII: Big Data and Privacy in Government

VII-1: Privacy Laws

There is a complex matrix of laws, regulations and practices that arise from the possible use of Big Data in Government, and might affect its usage. The major international, national, and provincial laws are summarized in Appendix XI-3~~XI-3~~ However, many other sector specific privacy laws and considerations exist that may also come into play, depending on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the type of consent obtained from the data subject, etc. The following observations can be made about the legislation:

Formatted: Hidden

- There is no single law or practice governing data privacy; legislation exists at all levels of government creating a complex matrix of international, national and provincial laws that govern the use of personal information in Canada and abroad.
- Although similar in concept, privacy laws are not always aligned; some laws also have cross-border and extraterritorial reach.
- Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA.
- Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may be in conflict with the Privacy laws.
- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes, and may therefore be difficult to apply.
- There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.

VII-2: Recent and Pending Changes to Privacy Legislation

All governments are struggling with ways to keep their data privacy legislation current, relevant, and usable in light of the rapid technological developments. Of particular concern are the new analytical tools that have the ability to mine data and analyze the ever-increasing data sources, and especially those that target personal information. Perhaps of even greater concern is the trend toward consolidation of existing databases into Big Data sources. The concentration of personal information from various sources adds complexity and risk. Privacy laws in the

international community are far from static, and changes are likely to have an impact on Canadian laws and practices as these occur.

"As business systems and processes become increasingly complex and sophisticated, organizations are collecting growing amounts of personal information. As a result, personal information is vulnerable to a variety of risks, including loss, misuse, unauthorized access and unauthorized disclosure. Those vulnerabilities raise concerns for organizations, governments and the public in general." (AICPA/CICA).

Recent changes made to legislation could have significant implications for personal data privacy and the rights of Canadians. The specific aspects of these laws are presented in Appendix XI-~~4~~XI-4. The laws include:

Formatted: Hidden

- Bill S-4 The digital Privacy Act
- Bill C-13 Protecting Canadians from Online Crime Act
- Bill C-51 Investigative Powers for the 21st Century Act (aka the "Anti-Terrorism Act")

VII-3: Legislative Restrictions, Guidelines and Safeguards Regarding Government Use of Personal Information

An increasing amount of information is available from the Canadian Government through its "Open Government" and its other initiatives; this trend is likely to continue. At the same time, controls have been established to try to ensure that personal information is only made available to those who are authorized to access it.

ACCESS TO INFORMATION AND PRIVACY PROGRAM (ATIP)

Systems are controlled and information is subject to review under the requirements of the Access to Information Act and the Privacy Act before being released. The ATIP program also permits citizens to determine what information government holds about them, and provides them with the ability to correct the information if it is inaccurate.

Government procedures also exist surrounding the handling of personal information by its departments and agencies. Guidelines issued by the Treasury Board include a Directive requiring the performance of an extensive Privacy Impact Assessment (PIA) before implementing or changing government systems, or altering the manner in which they process information. The PIA includes guidelines for the assessment of privacy implications before entering into contracts or making outsourcing decisions.

THE STATISTICS ACT

The Statistics Act permits StatsCan to enter into contractual agreements to share confidential information with other government departments under specific conditions:

1) Information can be shared with the statistical agencies of provinces and territories for statistical purposes if:
    a. The data subjects were notified at the time of data collection;
    b. The provincial agency has the authority to collect the information on its own; and
    c. The agency's confidentiality protection requirements are substantially the same as those of Statistics Canada.
2) Where information is collected jointly by Statistics Canada and any federal and provincial government department, municipal government or other incorporated body such as an association or university, and where data subjects are notified in advance of intention to share the data, and are given the opportunity at the time of data collection to refuse to allow their information to be shared.

The OPC has also highlighted the existence of other laws that supplement, but do not necessarily supersede, the Privacy Act and PIPEDA and which provide Canadians with additional protections for their personal information:

"Several federal and provincial sector-specific laws include provisions dealing with the protection of personal information. The federal *Bank Act*, for example, contains provisions regulating the use and disclosure of personal financial information by federally regulated financial institutions.

Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals." Source (Office of the Privacy Commissioner of Canada)

Therefore, many substantial controls do exist over the internal use of personal information by government departments and agencies.

VII-4: Implications for The Department of Justice

Determining which jurisdiction governs personal information is becoming much more complicated as information is gathered and/or transferred across legal jurisdictions and co-mingled in Big Data stores or linked with other information sources. It is also easy to lose track of the origin of the data over time, and especially if the organization operates across Canada or captures information on the Internet. Maintaining data accuracy and responding to citizen's

information requests becomes problematic. Courts around the world are struggling with data ownership and the determination of which laws will apply.

Most of the Big Data that is to be used by JUS is likely to consist of legal precedents and opinions, or possibly large quantities of evidence submitted in a court case to be analyzed using eDiscovery tools. Therefore, although there may be a few exceptions, (e.g. criminal records), JUS appears unlikely to capture and use a significant amount of personal Big Data, other than where it uses personal information contained in other government databases (e.g. StatsCan) for analytical purposes, and usually in aggregated form. However, the department may use personal information of its own staff members to assess efficiency and productivity of the various department functions. PIPEDA may also apply to aspects of litigation proceedings, depending on the context, when personal information captured in connection with litigation involves commercial organizations or is carried out in the course of commercial activities.

Regardless, JUS is likely to be involved in legal actions or discussions surrounding the use of personal data by other Government departments, and some of the evidence that it collects which includes sensitive or other personal information must be kept private. In these cases, JUS lawyers will need to be respect their obligations under PIPEDA by ensuring that any personal information collected, used or disclosed in connection with any anticipated or actual litigation (or any other use) needs to be done either with the consent of the individuals, or must otherwise meets one of the applicable exceptions to the knowledge and consent principles of PIPEDA or the Privacy Act.

Section VIII: Privacy Concerns - Big Data and Government

The issues raised by the establishment of Big Data sources are not necessarily new, but relate to the difficulty of managing and protecting such large banks of information. Also, the sheer volume of the data held by government – both collectively and about each individual – creates the concern that profiles of individual characteristics and behaviour can be established that are quite complete and accurate. The application of predictive analytics to that information could permit the prediction of future trends and behaviours of both societies and individuals. While this might have benefits, there is a darker side to the existence of mass stores of personal data if the capability was misused.

The main concerns, discussed further below, are likely to be in four broad areas:

- Big Data Management and Security;
- Individual Consent regarding permitted uses of the personal information;
- Transparency and government disclosure of how data is collected, stored and used; and
- Lack of trust in government.

VIII-1: Big Data Management and Security

Large electronic sources of personal information can have significant value to those with less honourable intents. Once accessed, huge amounts of information can be rapidly transferred and stored inexpensively and with relative ease, attracting theft for monetary gain or extortion where personal exposure might have adverse impacts for both individuals and governments. The more attractive the information, the greater the difficulty to protect against data breaches by sophisticated hacking communities or tools – both in state-sponsored or private hands.

The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. This could involve greater risk of misuse in the event of a data breach, and eventual misuse for identity theft or fraud. The risk to government and individuals must be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

The demonstrated ability of hackers to overcome the security of government websites (e.g. recent attacks by Anonymous on Canadian Government web sites) and the perception that personal information is at risk of being disclosed or used fraudulently undermines public confidence in the safety of having their personal information in government data repositories.

"Each of the Canadian Privacy Statutes contains safeguarding provisions designed to protect personal information. In essence, these provisions require organizations to take reasonable technical, physical and administrative measures to protect personal information against loss or

theft, unauthorized access, disclosure, copying, use, modification or destruction. These laws do not generally mandate specific technical requirements for the safeguarding of personal information." (Piper, 2015)

Somewhat surprisingly, there are no prescribed standards for implementing security controls to protect personal information; rather it is left up to organizations to use their own judgement to determine what is appropriate. PIPEDA and the B.C. and Alberta privacy acts only "require organizations to take reasonable steps to safeguard the personal information in their custody or control from such risks as unauthorized access, collection, use, disclosure, copying, modification, disposal or destruction." (Office of the Privacy Commissioner of Canada, n.d.)

Reasonable safeguards include several layers of security, including, but not limited to risk management; security policies; human resources, physical and technical security; and business continuity management. The reasonableness of security arrangements adopted by an organization must be evaluated in light of a risk assessment including a number of factors, such as the sensitivity of the personal information; the foreseeable risks; the likelihood of damage occurring and the resulting harm caused; the medium and format of the storage method, and the cost of putting preventative measures in place.

VIII-2: Consent to Use Personal Information

The so-called "secondary use" of personal data - i.e. the use of data that has been provided for one purpose for other purposes - is a growing problem in the digital world, and in the Big Data world in particular. There is also a grey area between what information might require explicit or implicit consent for its use. The rules surrounding the requirement for consent and the use of personal information is clearly laid out for the private sector in PIPEDA, but Big Data will create broader issues for the public sector as well.

In the past, some of this data was considered to have been provided with the individual's implicit consent that it would be used in accordance with disclosures made by organizations. However, legislation covering the collection of most personal data collected by private organizations in Canada now requires explicit consent for use in accordance with specific terms. Any proposed secondary use for other purposes isn't generally permitted unless the use is disclosed at the time of collections. This is especially true in situations regarding the use of one of the sensitive categories of information (see Appendix XI-6~~XI-6~~).

Formatted: Hidden

Subject to legal interpretation, The Privacy Act might provide the government with more flexibility in its use of information provided to its various departments in the normal course of business, including the sharing and exchange of this information between government departments in the form of a Big Data repository, so long as the information is adequately protected from improper access or uses. Such use is already being made for research purposes

(e.g. by StatsCan). Sections 7 & 8 of the Privacy Act appear to cover this use. However, in the future expansion of Open Data and Big Data, where information is spreading out in many directions, it might be more difficult to determine whether information is being used in ways that don't require some form of additional consent or opt-out capability, and there may be unintended consequences. The standard form of consent or notification provided by the government will probably have to be worded very carefully at the front end of the process, and the back end of the process will require some form of careful review to ensure that the information is not being used outside of legal boundaries.

VIII-3: Transparency and Government Disclosure

The 2014 OPC survey reported, "The vast majority (89%) of those who had heard something about government surveillance activities agreed that surveillance or intelligence gathering agencies should have to explain their activities to Canadians." (Phoenix Strategic Projections Inc, 2014)

In 2000, the Canadian Government began to create its first Big Data repository, which became known as "Big Brother". The database included information on the addresses, education, marital status and ethnic origin of Canadians. It also tracked a person's employment and social assistance history, and their income tax records. Plans to implement the database were shelved at the time due to concerns expressed by the OPC and in Parliament, and also because of the volume of public requests to see their personal information contained in the database. (CBC News, 2000)

The concerns of the Canadian public in this area remain today. A conclusion of the 2014 OPC survey was that "The majority of Canadians are not confidant that they have enough information to know how new technologies might affect their personal privacy." This would likely extend to the enhanced use of Big Data by government. "Canadians expressed varying levels of comfort with different ways in which government departments and agencies, including intelligence gathering organizations, could collect or share their personal information." (Phoenix Strategic Projections Inc, 2014)

Only about half of the OPC survey respondents felt that:

- They had a good understanding of what the Government did with personal information that it collects;
- They were confident that the government would take their concerns about handling of their data seriously;
- They were confident that personal information shared with government would not be misused, lost or stolen.

VIII-4: Data Breaches and Trust in Government

While there are no statistics regarding trust in the Canadian government to protect personal information, numerous highly publicized data breaches have occurred in Canada over the past few years, and the numbers have grown substantially:

"The federal government reported breaching the privacy of individuals more than 5,000 times last year — an all-time high, according to new figures. The data are only for six departments, so the 5,237 privacy breaches they reported in 2014 are likely just a glimpse at what happened across government. Even so, the figure is almost as many as had been reported in the previous 11-year period, including instances where a taxpayer's or organization's information was incorrectly released, lost or compromised." (Press, 2015). Public awareness of attacks on government has increased too with the recent highly publicized attacks on Government web sites by "hacktivist" groups, such as Anonymous.

Canadians are waking up to the possible uses of their personal information by government agencies. The December 2014 OPC survey found that "56% of Canadians have some awareness of surveillance and intelligence gathering activities." "Roughly half (49%) of Canadians have seen, read, or heard something about surveillance or intelligence gathering activities for the purposes of national security in the past year or so." (Phoenix Strategic Projections Inc, 2014)

The heightened awareness of Canadians is likely a result of the recent publicity of government surveillance and information sharing programs through public revelations by Richard Snowden and the debate surrounding Bill C-51 (now the Anti-Terrorism Act) and its potential implications for the privacy of personal information. 78% of those polled in the OPC survey said they were either very (44%) or somewhat (34%) concerned about law enforcement and security agencies collecting their personal information for government surveillance purposes.

VIII-5: Big Data Privacy Controls

While the use of Big Data and related technologies can create significant privacy concerns as highlighted above, some of the technologies available now also permit the implementation of sophisticated controls to protect individual rights of citizens by regulating how Big Data technologies are used. Examples of these controls include:

- Use of methods for the "tagging" of data to ensure use is restricted to the purposes for which it was collected or generated;
- Implementing purpose-based or user-based controls according to the permissions and restrictions established for this data, including access controls;
- Tracking user access to data and the purposes for which it is used;

- Implementing algorithms that provide alerts regarding inappropriate access and possible uses.

While the use of specific information by JUS may not raise broad privacy concerns, other information available to government agencies may cause issues when data is aggregated or concentrated in electronic form, and especially when data is merged from multiple agencies. Regardless, there can be great benefits to merging and analyzing this information, such as:

- For research and public policy development regarding health, social, economic, national statistical trends;
- To demonstrate transparency and accountability of government; and
- To achieve public participation through engagement.

There is tremendous value in having broader access to this information for research, analysis, and policy development. Big Data is being used by the US Department of Justice to analyze medical billing records to detect Medicare fraud, and they are looking at similar Big Data sources for the detection of other frauds. (Scannell, 2015). The increased sharing of information across government departments also creates complex relationships and can result in difficulties surrounding disclosure and transparency about the use of the information. One such example is the Canadian Open Government Portal that is intended to provide "greater transparency and accountability, increase citizen engagement, and drive innovation and economic opportunities through Open Data, Open Information, and Open Dialogue" (Government of Canada, n.d.).

Achieving full openness while maintaining appropriate controls over data privacy may be mutually exclusive objectives requiring some compromises. Legal privacy objectives can often be achieved through "de-identification" or "anonymization" of data, but the more heavily data is neutralized in this manner, the less useful it can become. In addition to legal requirements for compliance, there are also ethical considerations and, while privacy and confidentiality are somewhat different concepts, contractual and other agreements regarding the possible use of information (e.g. copyright) may need to be considered.

Focus groups during the 3rd International Open Data Conference held recently in Ottawa identified several privacy concerns and issues around open data:

- The public sector collects a great deal of sensitive personal information. While individual sources of anonymized or de-identified information might not reveal the identity of a person, the use of multiple data points that link or connect to others may make it possible to connect or triangulate between unrelated data points, making it possible to identify individuals.
- The use of Census and national statistical information can be problematic, even if data is aggregated, since individuals can often be identified within small groups or communities. Locational data can sometimes involve the same risk as a personal identifier "key", such as a name or social insurance number.

- The potential to profile, target or discriminate against vulnerable people or groups might be possible through matching of open data sources with information gained from other private sources.

The concern exists that while government surveillance will be made easier for protection against terrorism and illegal acts. (Open Data Ottawa Privacy Conference Notes)

VIII-6: Forecasting Canadian Public Opinion on Privacy and Big Data in Government

According to a study conducted by the Office of the Privacy Commissioner (OPC) in December 2014, "Nine in ten Canadians expressed some level of concern about the protection of their privacy, with 34% saying they are extremely concerned (up from 25% in 2012)." Further, "Canadians increasingly feel that their ability to protect their personal information is diminishing. Seventy-three percent, the greatest proportion since tracking began, think they have less protection of their personal information in their daily lives than they did ten years ago." (Phoenix Strategic Projections Inc, 2014)

Canadians are therefore aware and concerned about the privacy of their personal information, and increasingly so. The primary focus of Canada's privacy programs has arguably been on the use of personal information in the private commercial sector, and the protection of this data through PIPEDA and its enforcement by the OPC. The same degree of knowledge or awareness of the Privacy Act and the permissions afforded by it to government doesn't seem to exist.

At the same time, recent legislative changes (see Appendix XI-4~~XI-4~~) and public revelations concerning clandestine government surveillance programs by Western governments, including Canada, have not likely helped to ease public concern. Some vocal members of the Canadian public, in particular, are questioning whether the extent to which the legislation is being implemented is commensurate with the need.

Formatted: Hidden

Anti-Terrorism Bill C-51, in particular, appears to be the subject of much concern. Daniel Therrien, the Privacy Commissioner of Canada, is responsible for the independent oversight of Canada's privacy laws and compliance. He recently submitted an article published in the Globe and Mail in which he said:

"In my view, Bill C-51, in its current form, would fail to provide Canadians with what they want and expect: legislation that protects both their safety and their privacy. As proposed, it does not strike the right balance.

The scale of information-sharing between government departments and agencies proposed in this bill is unprecedented. The new powers that would be created are excessive and the privacy safeguards proposed are seriously deficient." (Therrien, 2015)

The focus on the use of Big Data to track people and groups casts a negative image. The Commissioner's comments and position on information sharing are likely to create further debate and shape public opinion regarding government Big Data and data sharing between departments and with other governments. As sharing of Big Data information by government agencies becomes more commonplace, Canadians may become increasingly concerned about the possible uses, and react negatively.

Addressing the lack of awareness by Canadians of the way in which their personal information is being used may require greater emphasis on public disclosure to help reduce concerns. One recommendation made by the recent report to US President Obama suggests the implementation of a Consumer Privacy Bill of Rights based on the Fair Information Practice Principles. While this approach might help confidence in the private sector, a broader "Citizen's Bill of Rights" might be more appropriate to help renew the trust in government to protect personal information in the face of the expanded use of Big Data. One of the Key Informants, Howard Deane, from the Consumers Council of Canada, expressed the view the level of trust might be elevated if the government was more transparent regarding how personal information that makes its way into their hands will be used (i.e. limits on use), and what protections will be put into place around Big Data to avoid its misuse.

There are also ethical and moral questions about how Big Data might be used by government, or disclosed to others for possible misuse. There is a difference between government predicting and disclosing broad statistics about crime and cancer rates on a macro scale and using the data to focus in on individuals. The more granular the information becomes, the more organizations might be tempted to use the information in negative ways.

In his Globe and Mail article, the Privacy Commissioner indicated that the new legislation would "provide 17 federal government agencies with almost limitless powers to monitor and profile ordinary Canadians, with a view to identifying security threats among them. The end result is that national security agencies would potentially be aware of all interactions all Canadians have with their government. That would include, for example, a person's tax information and details about a person's business and vacation travel." (Therrien, 2015)

Public opinion is often difficult to predict because it often varies by culture, and is subject to "trigger" events that cause rapid shifts - e.g. The Edward Snowden disclosures surrounding government surveillance involved such a shift. Other than the OPC survey conducted by Phoenix Strategic Projections Inc., there appear to be few detailed Canadian surveys and public opinion polls that specifically address this topic in detail, but recent studies done on public

perceptions and opinions in the US and EU confirm that people believe that the privacy and security of their personal information is at risk, as is their ability to keep their information confidential in such an open world. However, there are a number of recent international studies that support this view.[4]

A Welcome Trust study in the UK found that focus group participants distinguished between acceptable types of government uses of personal data according to the following factors:

- The Government identifying needs, planning resources and services, and allocating funds;
- Prevention and detection of crime and, including terrorism;
- Identifying social/population trends and statistics;
- Unearthing dishonesty (e.g. fraudulent benefit claimants and tradesmen)

While there was a general awareness of data collection by both government agencies and companies generally, the Welcome study found that the public views of the collection and use of personal data could be summarized as follows:

- The public consider the collection and use of personal data to be a big issue;
- When asked, the public are ostensibly opposed to any form of data use and collection by government and companies;
- In practice, the public consider there to be no alternative to sharing personal information with government and companies in the modern world and expect this to increase in future;

A significant proportion of the public expected to feel less comfortable about sharing personal data in future.

VIII-7:Summary

The 2014 study commissioned by the President of the United States regarding Big Data and Privacy included the conclusion that:

"Although the use of Big Data technologies by the government raises profound issues of how government power should be regulated, Big Data technologies also hold within them solutions that can enhance accountability, privacy, and the rights of citizens." "Responsibly employed, Big Data could lead to an aggregate increase in actual protections for the civil liberties and civil

[4] - PEW Research Study - Public Perceptions of Privacy and Security in the Post-Snowden Era - November 2014
- White House Study - Big Data and Privacy Review - May 2014
- EU Byte Study - Report on public perceptions and social impacts relevant to Big Data - March 2014
- Eurobarometer Report - Attitudes on Data Protection and Electronic Identity in the EU - June 2011

rights afforded of citizens, as well as drive transformation improvements in the provision of public services. " (Report to the Executive Office of the President)

It remains to be seen how the use of Big Data will translate into privacy concerns and the reaction by Canadians to the use of their personal information in Big Data repositories going forward. The level of trust in government, along with knowledge of why data is being collected and how it will be used also appear to be significant Issues, judging from recent public reaction to Bill C-51. There will likely be a need for programs to educate the public about these uses, and to promote the benefits, in order to establish a level of confidence and trust in the process, and to prevent a negative backlash such as occurred with the "Big Brother" database proposal in 2000.

Section IX: Major Findings and Conclusions

While Big Data is reforming many aspects of the world in which we live, the earliest successful models have been built on large databases of structured quantitative data, because this type of information is more easily and readily interpreted by binary computer logic. Although there are some exceptions, much of the information generated by the legal community or used in trials is unstructured data – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex and will take time to develop and refine.

Despite this, considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data elsewhere in the legal profession. The marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantages and cost efficiencies to those who adopt them.

Q1. Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

IX-1: Possible Big Data Strategy

JUS is in competition with other organizations in the legal community who will be making investments in these new technologies, and the Department will need to make similar investments, if only to be competitive and cost-effective as it conducts its business. It is involved in an "arms race", where all parties must move forward to avoid being placed at a strategic disadvantage. The strategic decision to be made by JUS is whether it should position itself as an early adopter of the technology, or be satisfied to be a "fast follower". The other decision will be how it should invest its limited resources to achieve its strategic objectives – i.e. what should the priorities be?

The one overarching conclusion that can be derived from the study is that large legal organizations that fail to plan for the implementation of these new technologies are likely to find themselves at a significant disadvantage from a competitive and cost-effectiveness standpoint. Donald Wochna, chief legal officer of Vestige Digital Investigations, was quoted in *Law Technology News* as saying: "Big Data in general, and predictive data analytics in particular, are the potential holy grail in the practice of law."

Q2. How can the Department of Justice adopt Big; Data for its own needs?

IX-2: Possible Uses of Big Data and Predictive Analytics in JUS

The possible uses of Big Data by JUS, and the implications thereof, include:

POSSIBLE APPLICATIONS OF BIG DATA BY JUS

The primary applications of Big Data analysis in the Department of Justice are expected to be the use of:

1) eDiscovery software tools to analyze and refine information in large databases of relevant documents for the production of evidence to be used in trials.
2) Predictive analytics and artificial intelligence to predict the outcome of cases based on a Big Data repository of precedents and legal opinions, which might ultimately be used to reduce time spent and effort devoted to settling cases or taking them through the trial process.
3) Data analytics techniques to analyze large databases of JUS operational statistics, with a view to improving individual performance and the overall productivity and cost-efficiency of the Department and, in future, to proactively position department resources to address emerging trends.
4) Data analytics to predict environmental trends, based on both internal (e.g. StatsCan) and external information (e.g. social media) that might be used to respond to the need for changes in government policies.
5) Automated tools for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process.
6) Automated tools to measure both individual performance and compliance with department and professional policies, procedures and standards and, in summary form, for management reporting of department performance and risk management.

SOME CONSIDERATIONS SURROUNDING BIG DATA IMPLEMENTATION

Lawyers who have already been exposed to the use of eDiscovery, predictive analytics and other advanced technology tools are recognizing some of the implications as well as the potential opportunities of working with advanced technologies and applying these tools to large repositories of relevant data. However, this is still a relatively new concept for many in the legal profession, and so it is difficult for them to know where to begin with plans for implementation. The following issues will need to be considered:

7) The legal world is definitely headed down a path where sophisticated technologies (e.g. Big Data and Predictive Analytics) will play an increasing role. Legal organizations that

fail to keep up will eventually find themselves to be at a competitive disadvantage in terms of managing litigation cost and achieving success in the trial process.

The implementation and adoption of complex new technologies can be a significant undertaking in large organizations such as JUS, and therefore takes considerable time (i.e. years) to accomplish. Advance planning is therefore critical to ensure that resources are available, and the implementation is a success.

8) Implementation of the new technology tools and processes will require a strong change management program, based on the inherent resistance of people to significant change. The input we received, both in JUS and externally, is that such a program is likely to be required in order to achieve widespread adoption of new technologies, and also systems that attempt to measure individual performance more closely.
9) Significant investment will be required over many years, in money and human resources to remain current with the external legal marketplace to avoid falling behind. This is especially true with respect to the use of Big Data and predictive analytics technology where there have been, and will be, significant developments in the legal community
10) JUS may not have access to the financial, human, and other resources required to move down all the emerging technology paths at once. The various options will need to be prioritized based on the projected cost/benefit before proceeding with any plans to implement Big Data, and considered as part of an overall departmental strategy.
11) Successful implementation is likely to depend on the ongoing commitment of JUS management to invest in the change, and to implement the tools required over a protracted period of time.

POSSIBLE COST EFFICIENCIES TO BE DERIVED FROM BIG DATA AND ADVANCED TECHNOLOGIES

The implementation of advanced technologies can be very expensive and disruptive to JUS, but the organization is likely to achieve both quality and cost-effectiveness improvements as a result. The following possible benefits were highlighted during our research:

12) Productivity and quality improvements would result advanced expert search technology and litigation support tools to better research information and relevant evidence;
13) Possible process and efficiency improvements could be achieved in JUS administration and operations;
14) Productivity could be improved through the use of advanced analytics to allocate litigation resources by predicting forward demand and adjusting supply of legal resources accordingly.
15) Costs might be reduced through the ability to predict case outcomes and resolve cases through the use of an alternate dispute resolution mechanism involving the use of Big Data and predictive analytics to achieve a settlement without going to trial.

16) Access to internal and external Big Data sources would permit better policy decisions.

Is JUS Positioned to Take Advantage of New Technologies?

While the main purpose of this study was to look forward at possible uses of Big Data in JUS, the current uses of Information Technology in the Department was also reviewed. This is important as a starting point because the transition to the use of Big Data and predictive analytics in most large organizations relies on having a relatively strong base of technology on which to build. However, some technical and organizational restructuring may be required in order to move forward with more sophisticated technology programs.

JUS appears to have a variety of available technology tools, but there is some question as to how extensively these tools have been accepted and are being used by Department staff. In addition some of the key systems (e.g. iCase) are aging and in the process of being replaced with Government standard tools, although implementation is just beginning.

Regardless, the conclusion is that:

17) No serious technology impediments were identified that would prevent JUS from moving forward with Big Data projects.

Q3. Are there promising practices in other countries and departments worth emulating? Where and what are they?

Practices in Other Countries and Departments

Sections D, E, and F of the report go into considerable detail about findings in this regard. The findings were mixed. Although there are some promising developments in other countries or departments that could be followed up, or developments to be followed, there don't seem to be any "magic bullets" at this time. However, there appears to be steady progress, and suppliers of technology in this area are making considerable investments.

18) Carrying on a "watching brief" while preparing to move forward as a clearer path emerges might be an appropriate strategy for JUS.

Q5. What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms, incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms

IX-3: Government Big Data, Data Privacy and Public Opinion

PRIVACY LAWS

A complex matrix of laws, regulations and practices impact the possible use of Big Data in Government:

19) Canada is one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.
20) Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes.
21) There is no single law or practice governing data privacy; legislation that exists at all levels of government – a complex matrix of international, national and provincial laws exist that govern the use of personal information in the private and public sectors in Canada and abroad.
22) Although national laws are similar in concept, privacy laws are not always aligned; some also have cross-border and extraterritorial reach. Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.
23) Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may also be in conflict with the Privacy laws. Other laws and agreements must also be considered – e.g. copyright laws and contract laws may govern the use and disclosure of personal and other data.
24) Jurisdiction of data privacy laws and the determination of which applies depends on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the consent obtained from the data subject, etc.

Q4. What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

IX-4: Other Challenges to Privacy in a Big Data World

DATA SECURITY AND BREACHES

25. The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. Government programs to expand access to data through the Internet also create additional points of potential entry for breaches to occur.

While data contained in Big Data repositories is unlikely to be released in volume, the ability to access a wide range of views of various information sources through available portals, and potentially to use sophisticated search capabilities to retrieve information can be causes for concern if the appropriate level of security and controls aren't in place.

26. The risk to government and individuals will need to be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

PUBLIC OPINION AND REACTION TO GOVERNMENT BIG DATA

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become very concerned about their information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to sharing Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. There are a number of factors, in particular, that might trigger a negative public reaction or, conversely, steps might be taken to mitigate a negative reaction from occurring.

27. The implementation of a Big Data repository by government is likely to require greater government transparency about the way in which government handles personal information in Canada, and a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

Q6. What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

28. There will likely be a need to re-examine and revise the various laws affecting personal data privacy in Canada, and especially as government and other Big Data projects are brought on stream. In this regard, any changes to the legislation need to be forward thinking regarding emerging technologies (e.g. the laws need "to go where the puck is going to be" with privacy legislation, and not where the puck has been) otherwise laws will become quickly out-dated.

Individuals with legitimate access rights (e.g. government employees) who are able to download information can also be a source of concern if that information is lost or compromised. There are controls that can be put in place to partially guard against these sorts of occurrence, but they are generally expensive and cumbersome to implement.

Section X: References

AICPA/CICA. (n.d.). *http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/PRIVACY/GENERALLYACCEPTEDPRIVACYPRINCIPLES/Pages/default.aspx.*

Armah, N. A. (2013). *Big Data Analysis: The Next Frontier.* Bank of Canada.

Braddell Brothers. (2015). *Singapore Litigation Procedure.* Retrieved from Braddell Brothers: http://braddellbrothers.com/litigation.html

Brown&Ehrenreich. (2015, July 13). Can Big Data and Privacy Coexist?

CBC News. (2000). *Ottawa breaks up 'Big Brother' database.*

Dwoskin, E. (2014, August 22). Can Big Data Improve Medical Diagnoses? *Wall Street Journal.*

EDRM. (2015, 1 1). *www.edrm.net.* Retrieved 6 9, 2015, from EDRM.net: www.edrm.net

Gartner Group. (2014). *Magic Quadrant for E-discover Software.* Gartner Group.

Gartner Inc. (2015, June 14). *IT Glossary.* Retrieved from Gartner Group: http://www.gartner.com/it-glossary/big-data

Google. (n.d.). *Google flu trends.* Retrieved from goog.org flu trends: http://www.google.org/flutrends/

Government of Canada. (n.d.). Retrieved from Canadian Open Government Portal.

IBM. (2015, June 16). *What is Big Data.* Retrieved from Big Data at the Speed of Business: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Joh, E. E. (2014, February). *Policing By Numbers: Big Data and the Fourth Amendment.* Retrieved from Washington Law Review: SSRN: http://ssrn.com/abstract=2403028

Krasnyansky, A. (2015, January 29). *Meet Ross, the IBM Watson-Powered Lawyer.* Retrieved from PFSK Labs: http://www.psfk.com/2015/01/ross-ibm-watson-powered-lawyer-legal-research.html

Leopold, G. (2014). AG Says Big Data Can Reform Sentencing Rules. *HPC Wire.*

Letouze, E. (2012). *Big Data for Development: Challenges and Opportunities.* New York: UN Global Pulse.

Library and Archives Canada. (n.d.). Legislative Restrictions: Records of the Government of Canada.

Library of Parliament Research Publications. (2014). *Lgislative Summary of Bill S-4.* (L. o. Parliament, Producer) Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?ls=s4&Parl=41&Ses=2&source=library_prb&Language=E#a1

Library of Parliament Research Publications. (2015). *Legislative Summary of Bill C-51: Investigative Powers for the 21st Century Act.* Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?Language=E&ls=c51&Parl=40&Ses=3&source=library_prb

Microsoft acquires Equivio. (2015, January 20). Retrieved from blogs.microsoft.com: http://blogs.microsoft.com/blog/2015/01/20/microsoft-acquires-equivio-provider-machine-learning-powered-compliance-solutions/

Office of the Privacy Commissioner of Canada. (n.d.). *A Privacy Handbook for Lawyers: PIPEDA and Your Practice.* Government of Canada, Office of the Privacy Commissioner.

Office of the Privacy Commissioner of Canada. (n.d.). *https://www.priv.gc.ca/resource/fs-fi/02_05_d_15_e.asp.*

Office of the Privacy Commissioner of Canada. (n.d.). *Securing Personal Information: A Self-Assessment Tool for Organizations.*

Open Data Ottawa Privacy Conference Notes. (n.d.).

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.* Rand Corporation.

Phoenix Strategic Projections Inc. (2014). *2014 Survey of Canadians on Privacy.* Canadian Federal Government, Office of the Privacy Commissioner.

Piper, D. (2015). *Data Protection Laws of the World.*

President's Council of Advisors on Science and Technology. (2014). *Report to the President: Big Data and Privacy: a Technological Perspective.* Washington, DC: Executive Office of the President.

Press, J. (2015, March 22). Federal government privacy breaches soar to record high. *Ottawa Citizen.* Ottawa, Ontario, Canada.

Report to the Executive Office of the President. (n.d.). *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.*

Scannell, K. (2015, January 12). DoJ uses big data to crack Medicare fraud schemes. *FT.COM.*

Shaw, J. (2014, April). Why "Big Data" Is a Big Deal. *Harvard Magazine*.

Solomon, H. (2013, June 27). *How Ontario faces big data privacy challenges*. Retrieved from IT World Canada: http://www.itworldcanada.com/article/how-ontario-faces-big-data-privacy-challenges/47722

Therrien, D. P. (2015, March 21). *Without big changes, Bill C-51 means big data*. Retrieved July 2015, from Globe and Mail: http://www.theglobeandmail.com/globe-debate/without-big-changes-bill-c-51-means-big-data/article23320329/

Transparency Market Research. (2014). *eDiscovery Market Global Industry Analysis, Trends and Forecast 2014 - 2020*. Transparency Market Research.

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. University of St Andrews, UK.

Section XI: Appendices

XI-1: Current Industry Leaders in eDiscovery (Gartner Group, 2014)

- kCura markets the Relativity platform that supports collection, legal hold, processing, review, analysis and production of evidence. Relativity is sold through a wide range of service providers and hosting partners, and through a growing direct sales channel.
- FTI Technology, a separate business unit of FTI Consulting, offers both e-discovery software and services. Its main Ringtail platform performs functions from processing to evidence production. The Attenex product, also offered by FTI, provides a combination of machine learning and visual graphics for ease of document review.
- Recommind is known for its predictive coding technology, and supports all stages of the EDRM. Axcelerate eDiscovery can perform legal hold, collection, processing, review, analysis and production of documents, with Early Case Assessment and predictive coding capabilities.
- ZyLAB has an integrated solution supporting all stages of the EDRM. ZyLAB Intelligent Information Governance is used for file analysis and classification. Its e-discovery technology architecture is horizontally scalable and can handle large datasets.
- HP's Autonomy eDiscovery tool supports the full process of EDRM. Their self-service eDiscovery OnDemand model is part of an ongoing product development initiative that addresses the market shift toward organizations that want to bring e-discovery in-house. The product has a wide range of stakeholders ranging from IT users to in-house general counsel.
- Nuix's products include eDiscovery, Enterprise Collection Center, Web Review & Analytics, and Legal Hold. Its technology also extends to other related use cases, such as archive migrations, information governance and information security.
- Exterro provides products to support e-discovery from identification through review. Its primary offering is the Exterro Fusion E-Discovery software suite, which is built on a single open platform. Exterro's Fusion Integration Hub allows integration of existing legal, e-discovery and other information management systems.

XI-2: International Privacy Legislation

The original concept of data privacy was developed long before the explosion in the use of information technology could be envisioned. The impact of the new technologies used both in personal lives and in business is now apparent. The use of technology to access and manipulate personal data will continue, and is placing serious pressure on existing data privacy laws and practices around the world to keep up with the pace of change.

Political, geographical and cultural issues have made it difficult to adopt a single standard set of laws for data protection. Many different laws and regulations prescribe the privacy and treatment of personal information processed in Canada and in other legal jurisdictions. Most data privacy regimes include a range of seven to ten common principles. Those with a fewer number generally combine some of the principles, with a result that is largely the same.

The major forms of international legislation in place that prescribe the treatment of personal information from a data privacy standpoint are:

EU Privacy Directive

One of the original, and arguably the strongest of the international privacy regimes, is the EU Directive (Directive 95/46/EC of the European Parliament and Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data) enacted by the European Parliament in October 1995. The EU Directive forms the basis for most national data privacy regimes in place around the world today. Only countries with privacy regimes in place that are deemed adequate by the EU are permitted to receive personal information from EU countries. The Canadian public sector Privacy Act and private sector "Personal Information Protection and Electronic Documents Act" (PIPEDA) and their application have made Canada one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.

OECD Guidelines

The OECD Guidelines, issued in 1980 and revised in 2013, prescribe eight Basic Principles for National Application that align broadly with the EU Directive. The ten PIPEDA principles largely conform to the OECD standards and principles. The OECD principles follow:

1. Collection Limitation Principle

There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

2. Data Quality Principle
Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

3. Purpose Specification Principle
The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

4. Use Limitation Principle
Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except:

a) With the consent of the data subject; or

b) By the authority of law.

5. Security Safeguards Principle
Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.

6. Openness Principle
There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

7. Individual Participation Principle
An individual should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;

b) to have communicated to him, data relating to him:

i) Within a reasonable time;
ii) at a charge, if any, that is not excessive;
iii) in a reasonable manner; and
iv) in a form that is readily intelligible to him;

c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and

d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.

8. Accountability Principle
A data controller should be accountable for complying with measures which give effect to the principles stated above.

APEC PRIVACY FRAMEWORK

The Asia Pacific Economic Cooperation (APEC) Privacy Framework provides for a flexible approach to information Privacy protection across member economies to avoid the creation of unrealistic barriers to information flows. The framework contains nine principles that are similar to the EU Directive, OECD Principles and PIPEDA. Privacy enforcement relies on authorities from participating APEC economies, including the Office of the Privacy Commissioner in Canada.

XI-3: Canada's Public and Private Sector Privacy Legislation

Canadian privacy legislation is aimed separately at the private and public sectors, and there are important distinctions between the two: the private sector includes privately owned, non-government entities, while the public sector includes organizations that are owned and operated by the federal, provincial, and municipal governments. Some examples are:

- Educational institutions such as universities, colleges, technical institutes, school boards;
- Provincial and regional health care institutions, nursing home operators, hospital boards and subsidiary health corporations;
- Local governments, including municipalities, police services and libraries.

Perhaps the most important difference between PIPEDA and the Privacy Act is that the former provides certain guarantees regarding the collection and use of personal information collected by private sector organizations, setting the stage for possible legal remedies and actions in the event of improper use and/or disclosure, whereas the Privacy Act does not set the same limitations on use of the information, nor does it provide for specific actions or remedies by government in the case of misuse, disclosure or data breach.

PUBLIC SECTOR PRIVACY LAWS - THE FEDERAL GOVERNMENT PRIVACY ACT

The Federal Privacy Act, first enacted on July 1,1983, applies to all of the personal information that the federal government collects, uses and may disclose about individuals or federal employees, i.e. it sets out policy surrounding the Government's collection, use and disclosure of their personal information in the course of providing services (e.g., passports, pensions, taxes).

The Privacy Act also sets out how federally regulated public bodies can collect, use, and disclose personal information, as well as how individuals can ask to access and update their personal information Examples of federally regulated public bodies include the:

- Bank of Canada
- Canada Revenue Agency (CRA)
- Canadian Space Agency
- National Research Council Canada
- Statistics Canada
- Treasury Board of Canada

The Act also gives the federal government a wide range of powers surrounding their possible uses and disclosure of personal information, subject to certain controls. The ability to share personal information across government agencies appears to be facilitated by the provisions of the Privacy Act. For example, Section 8 of the Act provides for disclosure in accordance with legal agreements between federal government departments, provinces and territories, First

Nations councils, and foreign governments for the purpose of administering or enforcing laws. Recent legislation further enhances the capability of sharing personal information across government agencies. See Appendix XI-4~~XI-4~~)

Formatted: Hidden

The Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with the Privacy Act.

PRIVATE SECTOR PRIVACY LAWS - THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC ACT

Federal legislation governing personal data privacy in Canada is provided in the Personal Information Protection and Electronic Documents Act (PIPEDA), which establishes the manner in which private sector organizations can collect, use or disclose personal information while conducting commercial activities in Canada. It also applies to the personal information of the employees of federally regulated organizations, such as telecommunications companies, banks and airlines. However, PIPEDA does not apply to non-commercial organizations such as charities or not-for-profits or political parties and some non-commercial associations.

PIPEDA generally applies to:

- Private sector organizations carrying on business in Canada in the provinces or territories, when the personal information they collect, use or disclose crosses provincial or national borders (except for the handling of employee information).

- Federally-regulated organizations with commercial operations in Canada, such as airlines, banks, telephone or broadcasting companies, but including their handling of health information and employee information.

PIPEDA sets out the following ten principles, which are closely aligned with the EU Directive, OECD Principles, and the principles adopted by the CICA and AICPA as "Generally Accepted Privacy Principles" (GAPP) (Appendix XI-5~~XI-5~~). The following privacy concepts are covered in the PIPEDA principles:

Formatted: Hidden

1. Accountability;
2. Identifying purposes;
3. Consent;
4. Limiting collection;
5. Limiting use, disclosure and retention;
6. Accuracy;
7. Security safeguards;
8. Openness;
9. Individual access; and
10. Compliance.

As with the Privacy Act, the Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with PIPEDA.

ROLE OF THE OFFICE OF THE PRIVACY COMMISSIONER OF CANADA

The Office of the Privacy Commissioner of Canada (OPC) advocates for the fundamental privacy rights of individuals through the establishment of an appropriate regulatory framework, and through the provision of independent oversight and monitoring of the application of PIPEDA to the private sector and the personal information handling practices of federal government departments and agencies to ensure compliance with the public sector Privacy Act.

OPC also acts as an ombudsman, working independently to:

- Advise individuals, government, businesses, and Parliament on emerging privacy issues;
- Investigate complaints and make recommendations based on findings; and
- Conduct audits under the two federal privacy laws; and
- Promote awareness and understanding of the protection of personal information.

PROVINCIAL LEGISLATION

Every province and territory has its own public sector legislation and these provincial acts also apply to provincial government agencies. Alberta, British Columbia and Québec have privacy legislation that is considered "substantially similar" to PIPEDA, so that the provincial act can be applied to private-sector businesses that collect, use and disclose personal information while doing business in those provinces. Ontario, New Brunswick, and Newfoundland and Labrador also have their own health care privacy legislation that supersedes PIPEDA in this area.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

Although provincial privacy laws are similar to federal laws, some important differences exist. For example, certain provincial privacy laws (e.g. Alberta) have special consent and transparency legislation that applies to organizations and/or their service providers who permit access to or disclose personal information to locations outside Canada. If this includes public sector bodies, it could create a conflict where information about an individual resident in a province might be shared with other governments, (e.g. in the case of suspected criminal, terrorist, or other activity, but where a crime has not yet taken place or been proven). It might also create problems with the potential capture, storage, and cross-border sharing of Big Data.

Only three provinces in Canada – Alberta, British Columbia, and Quebec - have their own private sector privacy legislation that supersedes PIPEDA; all others must comply with PIPEDA.

Organizations in Alberta, British Columbia, and Quebec therefore need to be careful of complying with both their own private sector privacy legislation as well as PIPEDA.

Alberta, Saskatchewan, Manitoba, Ontario, New Brunswick, Newfoundland and Labrador and Nova Scotia5have each passed health information protection laws to deal with the collection, use and disclosure of personal health information by public and private sector health care providers. Alberta and British Columbia have also passed privacy laws that apply to employee information. Some of these laws might not be considered to be sufficiently compliant with PIPEDA to be deemed substantially similar. Therefore, in some cases PIPEDA may still apply.

Some provincial sector-specific laws include provisions dealing with the protection of personal information. Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals.

PIPEDA doesn't apply to an organization where it operates entirely within a province that has privacy laws deemed substantially similar to PIPEDA, unless the personal information crosses provincial or national borders.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

(Office of the Privacy Commissioner of Canada)

XI-4: Recent Changes and Other Applicable Privacy Legislation

Bill S-4 The Digital Privacy Act

Bill S-4 amends the *Personal Information Protection and Electronic Documents Act*,2 the federal private sector privacy law. It does this in several notable ways, including by:

- Permitting the disclosure of an individual's personal information without their knowledge or consent in certain circumstances;
- Requiring organizations to take various measures in cases of data security breaches;
- Creating offences for failure to comply with obligations related to data security breaches; and
- Enabling the Privacy Commissioner, in certain circumstances, to enter into compliance agreements with organizations. (Library of Parliament Research Publications, 2014)

Bill C-13 Protecting Canadians from Online Crime Act

Bill C-13 deals with:

- The offence of non-consensual distribution of intimate images;
- Offences committed by means of telecommunication; and
- One aspect of the area of law, generally referred to as "lawful access", an investigative technique used by law enforcement agencies and national security agencies involving intercepting private communications and seizing information where authorized by law.

Bill C-51 Investigative Powers for the 21st Century Act (aka the "Anti-Terrorism Act")

Bill C-51 takes into account new communications technologies and equips law enforcement agencies with new investigative tools adapted to computer crimes. The new investigative powers within the legislation give law enforcement agencies the ability to address organized crime and terrorism activities online by:

- Enabling police to identify all network nodes and jurisdictions involved in the transmission of data and the ability to trace the communications back to a suspect. This includes information on the routing, but does not include the content of a private communication;
- Requires a telecommunications service provider to retain data to prevent its loss or deletion while law enforcement agencies obtain a search warrant or production order;
- Makes it illegal to possess a computer virus for the purposes of committing an offence of mischief; and

Enhances international cooperation to help in investigating and prosecuting crimes that extend beyond Canada's borders. (Library of Parliament Research Publications, 2015)

XI-5: AICPA/CICA Privacy Guidelines

The Ten Generally Accepted Privacy Principles

The ten Generally Accepted Privacy Principles are:

1. Management. The entity defines, documents, communicates and assigns accountability for its privacy policies and procedures.

2. Notice. The entity provides notice about its privacy policies and procedures and identifies the purposes for which personal information is collected, used, retained and disclosed.

3. Choice and consent. The entity describes the choices available to the individual and obtains implicit or explicit consent with respect to the collection, use and disclosure of personal information.

4. Collection. The entity collects personal information only for the purposes identified in the notice.

5. Use, retention and disposal. The entity limits the use of personal information to the purposes identified in the notice and for which the individual has provided implicit or explicit consent. The entity retains personal information for only as long as necessary to fulfill the stated purposes or as required by law or regulation and thereafter appropriately disposes of such information.

6. Access. The entity provides individuals with access to their personal information for review and update.

7. Disclosure to third parties. The entity discloses personal information to third parties only for the purposes identified in the notice and with the implicit or explicit consent of the individual.

8. Security for privacy. The entity protects personal information against unauthorized access (both physical and logical).

9. Quality. The entity maintains accurate, complete and relevant personal information for the purposes identified in the notice.

10. Monitoring and enforcement. The entity monitors compliance with its privacy policies and procedures and has procedures to address privacy-related complaints and disputes.

(AICPA/CICA)

XI-6: Other Categories of Personal Information

Sensitive Categories of Personal Information

Some personal information is considered sensitive. Sensitive personal information generally requires an extra level of protection and a higher duty of care. For example, some jurisdictions may require explicit consent rather than implicit consent for the collection and use of sensitive information. Some laws and regulations define the following to be sensitive personal information:

- Information on medical or health conditions
- Financial information
- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Sexual preferences
- Information related to offenses or criminal convictions

Source - (AICPA/CICA)

Nonpersonal Information

Some information about or related to people cannot be associated with specific individuals. Such information is referred to as *nonpersonal information*. This includes statistical or summarized personal information for which the identity of the individual is unknown or linkage to the individual has been removed. In such cases, the individual's identity cannot be determined from the information that remains, because the information is deidentified or anonymized. Nonpersonal information ordinarily is not subject to privacy protection because it cannot be linked to an individual. However, some organizations may still have obligations over nonpersonal information due to other regulations and agreements (for example, clinical research and market research).

XI-7: List of Key Informants

The following key informants were interviewed as part of the research process for this study

- Ms. Marj Akerley
  Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Kelli Brooks
  Principal in Charge, Evidence and Discovery Management
  KPMG LLP
  3020 Old Ranch Parkway, Seal Beach, California, USA
  USA

- Mr. Richard Cumbley
  Partner, Information Management and Data Protection
  Linklaters LLP
  1 Silk Street, London, United Kingdom

- Mr. Howard Deane
  Chair – Emerging Information Technology Committee
  Consumers Council of Canada
  1920 Yonge Street, Toronto, Canada

- Mr. Toundjer Erman
  Director, Business Management Strategic Planning and Business Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Dera J. Nevin
  Director of eDiscovery Services
  Proskauer
  Eleven Times Square, New York, New York, USA

- Mr. Chris Paskach
  Managing Director
  The Claro Group
  350 S. Grand Ave., Los Angeles, California, USA

- Mr. Jean-Sébastien Rochon
  Deputy Director and Counsel,
  National Litigation Support Services/National eDiscovery and Litigation Support Services

- Mr. Dominique Roy
  Director, Business Applications
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Julie V. Roy
  Supervising Counsel, National Litigation Support Services/National eDiscovery and Litigation Support Services.

- Ms. Tracy Sampson
  Depute Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Dugald Topshee
  Director, Client Relationship Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Eric Ward
  Senior Counsel, Public Law Sector – Information law and Privacy Sector
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Dr. Anthony Wensley
  Associate Professor, Department of Management,
  University of Toronto Kaneff Centre, 3359 Mississauga RD N Mississauga, Canada

- Mr. Omid Yazdi
  Partner, Forensic Services
  KPMG LLP
  550 South Hope St. Los Angeles, California, USA

Tilhoff, Tanya

From: Svarckopf, Jennifer on behalf of IM-IT Governance / Gouvernance GI-TI
Sent: 2015-Jul-31 2:30 PM
To: Akerley, Marj; Beaman, Peter; Buffam, Jennifer; Candline, Karen; Champagne, Michel; Gervais, Michèle; Hammoud, Katie; Hudson, Michael; Jakob, David; Lipinski, Stan; Livingstone, Edward; Lyon, Carla; Marion, Yves; McIntyre, Janet; Noftle, Tracie; Oberoi, Michele; Platt, Diane; Poliquin, Stéphanie; Rochon, Jean-Sebastien; Sampson, Tracey; Shuttle, Paul; St-Jean, Timothy; Svarckopf, Jennifer; Thibault, Darlene; Topshee, Dugald; Wetter, Colin; Yazar, Inanc
Cc: Benabdeljalil, Asmaa; MacLean, Alyson; McDonald, Susan; Li, Ting
Subject: FW: For Comment: Draft Big Data

Le français suit.

Hello,

The Research and Statistics Division (RSD) has contracted with E.S. Tunis and Associates (ESTA) to investigate emerging trends affecting the current and possible future uses of Big Data and Privacy within the Department. This work was undertaken by RSD in response to direction from the Deputy Minister that the Division engage in forward-looking exercises to explore issues that may impact the Department in the future. The first phase of the project consisted of research gathering – both from authoritative literature and interviews of Key Informants (including Departmental officials from IT, Public Law, Litigation Branch, and Business Analytics). The contractors have provided a draft report on the first phase of the project and RSD is requesting comments from members of the Business Transformation Committee **by August 13th**. Please send your comments to **Alyson MacLean,** Acting Director, Research and Statistics Division **(amaclean@justice.gc.ca).** Thank you!

****

Bonjour,

La Division de la recherche et de la statistique (DRS) a passé un contrat avec E.S. Tunis and Associates (ESTA) pour effectuer une recherche sur les nouvelles tendances ayant une incidence sur les usages actuels et futurs des données massives et la protection des renseignements personnels au sein du Ministère. Ce projet de recherche a été entrepris par la DRS à la suite de directives fournies par le sous-ministre selon lesquelles la Division devrait mener des exercices prospectifs; ceux-ci viseraient à explorer des questions qui pourraient avoir une incidence sur le Ministère à l'avenir. La première phase du projet correspond à la collecte de données de recherche – tant dans des ouvrages faisant autorité que dans le cadre d'entrevues menées auprès d'informateurs clés (notamment les fonctionnaires du Ministère du domaine des TI, du Droit public, de la Direction du contentieux et de l'Analytique des affaires). Les entrepreneurs ont fourni un rapport provisoire sur la première phase et la DRS demande aux membres du Comité de la transformation des activités de fournir leurs commentaires **d'ici le 13 août**. Veuillez faire parvenir vos commentaires à **Alyson MacLean,** directrice intérimaire, Division de la recherche et de la statistique **(amaclean@justice.gc.ca)**. Merci!

Big Data and Privacy Implicati...

Possible Big Data Uses by the Department of Justice

And Related Privacy Concerns

Draft Version 2

Version Date: July 24th, 2015

Prepared By:

2B-268 FIRST AVENUE OTTAWA, ON CANADA K1S 2G8

T 613 594 3033, F 613 594 8928

info@estaconsulting.org

www.estaconsulting.org

Field Code Changed

Field Code Changed

Table of Contents

Section I: Executive Summary

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends". What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement.

Big Data is reforming many aspects of today's world, and organizations everywhere are finding ways to use it to achieve competitive advantage. The Canadian government will eventually be obliged to adopt Big Data applications in order to remain internationally competitive. An overall strategy, which considers all of the relevant issues, would help the Government of Canada and all of its departments and agencies to harness Big Data while fulfilling its responsibility to protect the public.

The Department of Justice (JUS) asked E.S. Tunis & Associates Inc. (ESTA) to conduct research into applications and uses of Big Data being made in legal and justice systems that might be considered for use by JUS, and what a Big Data strategy might look like for the department. ESTA was also requested to consider the potential data privacy and protection implications associated with the use of Big Data by the Department. The research method included a review of primary, and secondary sources both internal to JUS and external. Following is a summary of the research findings.

Big Data Applications in the Justice System

Considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data in the legal profession. A large part of the information generated by the legal community or used in court proceedings is in electronic form, but much of this is unstructured – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex; they have taken time to develop and refine, but they are now coming rapidly on stream. In fact, the marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantage and cost efficiencies to those who adopt them.

There is widespread and growing use in western countries of intelligent eDiscovery software tools to analyze and refine large files of relevant documents for the production of evidence to be used in trials. New sophisticated analytics programs are emerging in the US to accurately predict case outcomes without the need to go to trial. Other potential uses of Big Data applications for consideration by JUS might include:

- Techniques to enhance and analyze large operational databases such as the JUS Case Management and Timekeeping systems to improve JUS productivity and cost-efficiency and, in future, to manage resources to address emerging trends.

- The use of data analytics to predict environmental trends using internal (e.g. StatsCan) and external (e.g. social media) information to contribute to policy debates.

- The use of automated tools and Big Data sources for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process.

Promising innovation is also taking place in the justice systems of other countries. Singapore launched a countrywide Integrated Electronic Litigation System (iELS) for all litigation to optimize scheduling of court dates, streamline court filings, and provide case management manage high volume litigation. iELS is accessible from anywhere through an internet browser.

The development and adoption of a Big Data strategy by JUS will not be a simple or inexpensive undertaking. It would require careful planning and long-term commitment if the strategy is to be successful. It will not be possible for JUS to stand still in this area. JUS lawyers will find themselves at a competitive disadvantage to other lawyers in courtrooms, and these pressures will inevitably initiate change. The strategic decision to be made is whether JUS will be an early innovator or a "fast follower". Either way, a careful planning and budgeting exercise will need to be undertaken.

JUS already makes use of technology applications that will provide it with a strong base from which to move forward with the deployment of Big Data and predictive analytics systems. Over the long term, it is predicted that the implementation of Big Data applications will provide both quantitative and qualitative benefits to JUS.

Data Privacy Considerations

Almost by definition, the concept of Big Data in government runs contrary to the concepts of personal data privacy, because a Government Big Data repository must ultimately contain a great deal of personal information about its citizens. Even if a data source is carefully screened to ensure that data is appropriately "de-identified", these protections may disappear when the data is combined with other sources for other uses. This raises potential privacy concerns and, depending on the situation, the potential for negative public opinion.

The implementation of Big Data systems by JUS specifically, and the Canadian government generally, will be challenging from a number of different standpoints.

- There is no single law or practice governing data privacy across Canada; different laws govern the privacy of personal information in the Public and Private Sectors and legislation exists at all levels of government. Although many laws are similar in concept, they are not always aligned, leading to a complex matrix of legislation and practices that surround the use of personal information in the private and public sectors in Canada.

- Some privacy laws also have cross-border and extraterritorial reach. Canada is one of the few countries accepted by the European Union (EU) as having adequate data privacy protections for personal data transfers from the EU. While this status speaks to the strength of Canada's privacy laws, it needs to be preserved as it gives Canada an economic advantage over the many other trading nations that do not have the same status.

- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes. These laws may need change or, at a minimum, to be reconciled as to how they apply in practice.

PUBLIC OPINION ABOUT GOVERNMENT BIG DATA AND DATA PRIVACY

Canadian public opinion about government use of Big Data mainly surrounds how their information will be used and protected. Canadians will be concerned with the security and privacy of their information held by government:

- From an IT security standpoint
- From a transparency standpoint (having knowledge of what is being done with their data).
- From a trust in government standpoint.

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become concerned about their personal information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. The implementation of a Big Data repository by

government is likely to require greater government transparency about the way in which government handles personal information in Canada, and also a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

SUMMARY

There is probably no alternative to the future use of expanded Big Data applications and repositories by JUS. It will become an imperative, if the operation of the Department is to remain cost-effective and competitive. Ultimately, the privacy concerns that arise from the use of Big Data by JUS, on its own, are likely manageable. The implications of the increasing and much broader capture and use of Big Data by government in general creates a number of legal, policy and other issues that JUS will inevitably need to help to resolve as it moves forward with Big Data applications.

Section II: Introduction

II-1: The Evolution of Technology and Big Data

The explosion of computing, electronic sensing and digital communications technology in today's society has led to an exponential growth in online data; we create roughly 2.5 quintillion bytes of data a day, so much that an estimated 90% of the data currently in existence were created in the last two years (IBM, 2015). Large, growing subsets of this mountain of data are referred to as Big Data.

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends. In particular, these findings relate to human behavior and interactions. For the most part, these are datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Brown&Ehrenreich, 2015). What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement (Ward & Barker, 2013).

Big Data repositories are a result of the exponential increase in the amount of data being captured, combined with advances made in low cost digital storage media. Almost all transactions are now done online, and most documents and forms are now available in digital form only. Internet-enabled devices that are capable of capturing personal, environmental and geolocational data surround us. This data is being used and combined in increasingly innovative ways that were often not anticipated during the initial collection process. Governments and private organizations alike are beginning to recognize the value of this data, and are investing heavily to harvest it to gain a competitive edge and other strategic advantages.

As data repositories have expanded and evolved, so too have the methods and processes that permit data search and manipulation. The cost of storage has decreased to the point where much data is kept indefinitely, often because it is easier and cheaper to do so than to devote resources to culling it. In the meantime, advances in processing power and the creation of new ways to combine and analyze the data have permitted the combination and parsing of the data for novel uses.

This new flood of information has led to a large number of opportunities across a wide variety of sectors, while at the same time giving rise to some new privacy concerns as more data is gathered in a world where many electronic devices are now internet enabled, and are beginning to monitor and store information on almost everything that we do. Given enough data about an individual, it is possible to create a very detailed profile that removes all prospects of future privacy.

II-2: The Inherent Conflict between Big Data and Data Privacy

The data that exists in a Big Data world must ultimately include a great deal of information about real people. In the past, this information existed in siloes that were, for the most part, physically separated because information was stored in paper files. Even in the early days of electronic data processing, information was stored on devices that were only accessible by individual computers with no connection between them. In today's world of Big Data, these electronic files are capable of being linked both physically and logically together to permit broader information access and greater system functionality.

Clearly, there is growing value in harnessing Big Data. Predictive modelling using Big Data sources will permit doctors to make more accurate medical diagnoses (Dwoskin, 2014). Medical diagnostic programs may soon be capable of using Big Data findings to review a patient's entire medical history, X-Rays and results of medical tests online – from virtually any location in the world – to make a diagnosis. Vendors can use multiple sources of information to predict retail trends and match their supply of goods and resources with anticipated demand. Governments can monitor health and other emerging social trends in their countries to forecast the need for public programs, resource allocations and budgeting.

An individual's privacy has long been considered a fundamental human right. However, the Canadian Charter of Rights and Freedoms, when enacted in 1982, didn't anticipate a world where an individual's personal information could be captured and stored in such minute detail, nor the ways in which it might need to be specifically protected. Sections 7 and 8 of the Charter have often been interpreted to provide these protections, but may not provide the required degree of specificity in a world where the various permutations and combinations of the data make it very difficult to ensure individual anonymity.

One of the first big uses of analytics applied to Big Data sources in government has been by the intelligence community, which developed programs such as Carnivore[1] to monitor and analyze large amounts of electronic communications in order to detect subversive activities. Predictive analytics are also being used to forecast crime levels based on regional and local demographics. This information is also being used, primarily in the US, in predicting an offender's likelihood of reoffending as a basis for sentencing decisions. Big Data history is already being used to predict future population trends. As more data is captured about the everyday activities of individuals, it will not only be possible to make predictions about their health and welfare as a basis for improvement, but also whether they may be more susceptible to committing criminal acts before they commit them. The Big Brother world of George Orwell's "1984" might have arrived.

[1] Carnivore was a system implemented in the US in 1997 by the Federal Bureau of Investigations to monitor email and electronic communications sent over the Internet

While there are ways to disguise personal information in large data sources, the current focus of investment and research is mainly on ways to harvest Big Data, rather than on how to protect it, along with an individual's data privacy. An appropriate balance will need to be established if Big Data and privacy are to co-exist peacefully in Canadian society.

II-3: Why Big Data?

Regardless of the challenges, Big Data offers considerable opportunities to the Department of Justice (JUS). In order to take advantage of the opportunities, and to minimize the negative effects of Big Data, JUS needs to develop a clear picture of the current state and likely near-term evolution of the technology. To this end, JUS commissioned ESTA Consulting to conduct research into:

- The impact of Big Data in the context of current privacy laws in Canada;
- Ways in which JUS could adopt Big Data for its own needs;
- The implications/opportunities of Big Data including the possible role for JUS, and whether a "Big Data Strategy" would help; also more generally for the Government of Canada.

Implementing a Big Data strategy is not a simple task, especially for organizations the size of JUS or other federal public departments. All organizations now use Information Technology (IT) to a greater or lesser extent, but it is important for organizations to understand their current use of technology as a prerequisite for planning how they might move forward. The purpose of this report is therefore threefold:

1) To broadly review the current applications of IT by the Department of Justice in order to help it assess its position vis-à-vis other legal organizations with respect to the implementation of the advanced technologies and techniques employed by others to harness the power of Big Data in the public and private legal sectors;
2) To identify some of the privacy issues from a legal or regulatory standpoint that might stem from the availability and use of Big Data by JUS and the federal government, both currently and in the foreseeable future;
3) To consider the implementation of Big Data by the Government of Canada and some of the broader issues that could arise from this use, including public opinion and reaction.

Section III: Methodology

III-1: Project Scope

JUS requested the following scope in the form of questions to guide the direction of the research:

SECTION 1: HOW THE DEPARTMENT OF JUSTICE CAN MAXIMIZE THE USE OF BIG DATA?

Q1. Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

Q2. How can the Department of Justice adopt Big Data for its own needs?

Q3. Are there promising practices in other countries and departments worth emulating? Where and what are they?

SECTION 2: PRIVACY

Q4: What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

Q5. What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms, incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms.

Q6. What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

III-2: Research

The following research activities were undertaken:

INITIAL RESEARCH

Initial research was performed to assist with the scope of the research, and in planning. This involved an initial review of material available online, and meetings and discussions with JUS research staff members to clarify roles and responsibilities. Initial research also included a literature review to identify existing and near-future uses of Big Data in the legal sector, both

private and governmental. The literature review also helped to better define the scope of the research.

PRIMARY RESEARCH

Primary research consisted of interviews with Key Informants, both internal and external, to identify issues and help establish an accurate overview of the industry, as well as to assist in determining the criteria to be used in analyzing the results and developing the conclusions. Selected Key Informants' views were solicited to help shape some research, highlight background issues and subject matter, and provide some assistance with key observations and conclusions. Their comments, where relevant and notable, were included verbatim in the report. (See Appendix XI-7~~XI-7~~ for the complete list of Key Informants interviewed).

Formatted: Hidden

Key Informants were also asked for their views and observations on the subject of potential Big Data uses in JUS and the Federal Government, and on the associated data privacy issues and public perception, in order to identify the issues and to establish a general overview of the environment and the potential issues and concerns. A Key Informants plan and guide was assembled to direct discussions with the informants, but questions were modified for each interview to match the particular area of expertise of the Key Informants. The focus of the questions was on the direction of Big Data development in the legal marketplace, and possible data privacy implications associated with the use of Big Data, both by JUS, and more broadly by the Canadian Government.

In order to gain an understanding of Canada's use of Big Data in relation to the rest of the world, research was also conducted into the uptake of the identified technologies in various jurisdictions. An overview of other ways foreign governments use Big Data was also established.

SECONDARY RESEARCH

A broad background and view of the environment, drivers, issues, and industry players was developed from the initial and primary research. Published reports, research papers, websites, Internet sources on the topics, together with media reports, were then examined. Further research was then conducted into each of the identified Big Data uses in order to understand their capabilities, limitations and methods of use, and to identify the most common product options in use. Secondary research focused on areas of legal administration related to the business of JUS.

All research was conducted with a view to produce an initial identification of emerging issues and risks in the use of Big Data, primarily by JUS, but also more generally by government agencies.

Drafts and the final reports were reviewed with JUS staff to ensure accuracy and to verify scope coverage.

Section IV: The Emerging Uses of IT In the Field of Law

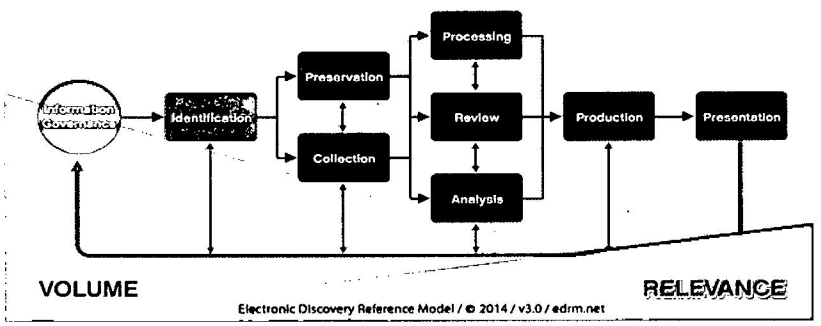
While technology applications, such as practice management systems (e.g. for time-keeping and financial management) have been used for some time in legal service organizations, the broader use of technology tools has been a relatively recent development. This has likely been driven in part by the explosion in the amount of unstructured (i.e. text-based) information in electronic form, and partly by innovations in the technology world to improve the ability to search, correlate and interpret this unstructured information in meaningful ways to gather and interpret evidence used in legal cases.

This section discusses some of the rapidly evolving uses of technology in the legal profession, including enhancements attributable to the emergence of Big Data; the sophisticated tools used to analyze large stores of data in the areas of eDiscovery and evidence gathering; legal research; the prediction of trial risk and outcomes and in the use of advanced data analytics for practice management and to achieve productivity improvements.

IV-1: eDiscovery Methodology and Tools

Electronic discovery (eDiscovery) tools include software designed and used to identify, preserve, collect, process, review, analyze and ultimately to produce information in electronic form to support the legal discovery process as legal cases are being conducted. eDiscovery software capabilities include the ability to identify, preserve, collect, process, review and produce information for use by counsel. These capabilities are generally conducted in a sequential order prescribed in the Electronic Discovery Reference Model (EDRM, 2015), a framework that has been established and is broadly accepted by eDiscovery practitioners.

Electronic Discovery Reference Model



During the initial phases of the eDiscovery process, the data collection and early assessment capabilities of eDiscovery software is used to refine the data so that an initial evaluation can be made regarding the information quality, the location of information that is available for use in a case, and what additional resources might be required for its effective evaluation. A risk assessment is generally performed during this stage to determine whether any restrictions might govern the use of the data, such as policies or data protection laws.

Subsequent phases of the eDiscovery process generally include technology assisted review tools that employ analytics-based machine learning technology. These use statistical techniques to "train" the software to review the electronic files, thus reducing the required amount of manual review to improve overall cost-effectiveness of the review process.

eDiscovery tools have evolved considerably since they were first introduced to the legal marketplace. In its May 2015 "Magic Quadrant for eDiscovery Software" study, Gartner Group (Gartner Group, 2014) studied 18 of the top organizations providing eDiscovery solutions and services to the marketplace today. They positioned the 7 organizations described in Appendix XI-1~~XI-1~~ as the current industry leaders.

Formatted: Hidden

Key Informant Kelli Brooks, who heads up KPMG's Evidence and Discovery Management Group in the US, noted that the kCura Relativity platform is the most commonly used tool, but indicated that the following eDiscovery platforms had potential for creating significant developments in the eDiscovery industry:

- Equivio is a relatively new Israeli text analysis start-up company that was bought by Microsoft in 2015. Industry speculation is that Microsoft plans to integrate the Equivio machine learning technology into Office 365 in future.
- Brainspace is a revolutionary new tool that can be used to reveal complex relationships between documents for review.

PREDICTED CHANGES IN THE EDISCOVERY MARKETPLACE

Transparency Market Research, a U.S.-based provider of syndicated research, customized research, and consulting services estimated that the Global eDiscovery market was valued at USD 5.56 billion in 2013. Government and regulatory agencies were the largest end-user segment in 2013, accounting for about 51% revenue share of the global eDiscovery market. They expected the market to grow at a cumulative annual rate of 15.5% from 2014 to 2020 as eDiscovery solutions find widespread applications in government and regulatory agencies, small, mid and large-sized enterprises and law firms. (Transparency Market Research, 2014)

The eDiscovery marketplace will also change as electronic evidence expands from the current analysis of email, documents and voice mail to include social media and mobile data. Increases

in data transfers between inter-connected business systems will require growth in the ability to analyze structured data. A combination of human skill and sophisticated software tools such as predictive coding and structured data analytics will be required to analyze these more complex evidence streams. A number of large players in the IT world are investing heavily in both eDiscovery software and tools for predictive analytics in the legal marketplace. These include:

HP Autonomy – The Hewlett-Packard purchase of the Autonomy search engine in 2011 for $10.3 billion set the stage for their entry into the eDiscovery marketplace, and they have continued to invest heavily since in new functionality (e.g. a cloud-based offering) to expand in the legal marketplace;

ROSS – a result of collaboration between the University of Toronto and IBM, using IBM's Watson Artificial Intelligence engine for legal research (Krasnyansky, 2015);

Microsoft's purchase of the rapidly growing Equivio in January, 2015 for a rumoured $200 million gave them access to "a provider of machine learning technologies for eDiscovery and information governance. We are making this acquisition to help our customers tackle the legal and compliance challenges inherent in managing large quantities of email and documents." (Microsoft acquires Equivio, 2015).

Key informant Dera Nevin advised that two important issues must be addressed before an organization can move forward with plans to capitalize on the use of Big Data for eDiscovery or more sophisticated applications (e.g. Artificial Intelligence (AI) and Predictive Analytics):

1) An appropriate information governance structure must be in place so that the organization has knowledge of what electronic information they have and where it is stored. In the past, legal organizations have been overly reliant on the use of paper documents, and a significant cultural change is required to overcome this issue.
2) Organizations need to standardize on a limited set of eDiscovery tools to permit legal counsel to become experienced with their use. Lawyers won't become experts in programming, but they will need to become adept in future at using sophisticated tools to search for and manipulate data.
   Although software standardization is a desired goal, Ms. Nevin observed that large legal organizations like the Department of Justice are also exposed to a wide variety of legal scenarios, and since there are specific strengths and weaknesses of the various tools on the market, a single eDiscovery solution might not be suitable for every case. The need to differentiate between structured and unstructured data may also require different tools.

STATUS OF THE USE OF EDISCOVERY TOOLS IN JUS

JUS IT representatives indicated that the Ringtail tool, from FTI Technology, is used across the Government for evidence management. In addition to JUS, the RCMP, PCO, PPSC, Election Canada, and the Treasury Board apparently use Ringtail. However, at least one of the key informants we spoke to expressed the view that JUS use of the tools was not as extensive as it might be, and that JUS might be lagging the private sector in this area.

Jean-Sébastien Rochon and Julie Roy of the National eDiscovery and Litigation Support Services group indicated that about 14-16 paralegal positions are devoted to the support of Ringtail and eDiscovery tools. Ringtail is only used for files involving more than 5000 documents as it is not cost-effective on smaller cases.

A problem highlighted with the current implementation of Ringtail is that documents from the 1700 cases stored in the system are held in "silos", so documents used for evidence in one case aren't available for use in others, although they might be useful. Going forward, Rochon and Roy hope to restructure the Ringtail database so that over 25 million pages of documents could be searched across the system and made available if they are relevant to other cases, and aren't subject to legal privilege.

Another problem they identified was that legal units assigned to other government departments sometimes use other eDiscovery tools not recommended by JUS. These create files that aren't compatible with JUS and therefore can't be shared.

USE OF EDISCOVERY IN OTHER JURISDICTIONS

A review of other jurisdictions finds mixed approaches to the application of eDiscovery. Table 1 (below) shows an overview of the use of eDiscovery and governing laws in various countries:

Table 1: EDiscovery Around the World

Country	eDiscovery Legislation	eDiscovery Use
Canada	• Sedona Canada Principles Addressing Electronic Discovery (1st ed 2008, 2nd ed 2015) (Federal, compatible with all provinces and territories except Quebec, based on US) • Ontario, Nova Scotia, Manitoba, Saskatchewan, Alberta and BC all have guidelines for eDiscovery based on the Sedona principles • Quebec, as a civil law province, has different rules	• Widespread • Following the American example

United States	• Legislation in effect since 2006 (meet and confer), updated 2007, 2015 (pending)	• Widespread • Pioneering and exporting eDiscovery to the world
United Kingdom	• Legislation in effect since 2009 (meet and confer), updated 2013 • Very specific eDiscovery guidelines and requirements • Litigation budget is required early in the process	• Widespread • Some jurisdictions require all cases to use eDiscovery, some allow the judge to make the decision on a case by case basis
Australia	• Legislation in effect since 2009 (meet and confer) • Update in judge training program includes managing eDiscovery and electronic case management	• A court can order all discovery for a case be done electronically. • Most courts have implemented individual guidelines specifically for eDiscovery
New Zealand	• Legislation in effect since 2012 (meet and confer) • All discovery is now electronic, unless the court decides otherwise.	• Waited a long time to make rules, and had a chance to see what other commonwealth countries did • EDiscovery is now ubiquitous
Japan	• No laws governing eDiscovery for domestic litigation.	• Not very common in non-governmental cases • An expectation of data production exists for government investigations • Slowly gaining popularity, mostly driven by international litigation, particularly with US law firms and vendors
Korea	• No specific eDiscovery laws • Very strict privacy laws, including a requirement that all corporate and personal data be hosted domestically	• Virtually non-existent
Singapore	• Integrated Electronic Litigation System (iELS) implemented in 2013	• All cases use eDiscovery through the iELS

IV-2: Timekeeping, Document and Case Management

JUS has used its proprietary iCase tool for a number of years to store documents used in litigation. iCase is also used for time and case management. The JUS IT group indicated during our interview that a major goal is to align JUS systems to the extent possible with prescribed federal government standards. GC Docs has been adopted as the standard for record keeping and document management, with Microsoft SharePoint 2013 as the front-end interface and

system portal. JUS will be converting, but implementation is only in the early stages, with migration of existing content occurring in the later stages.

IV-3: Legal Research

Internal and external information sources are used for legal research. *Justipedia* is the central legal knowledge management repository for the Department of Justice. It contains legal opinions, pleadings and facta, agreements and other precedents and tools. It is also used to access legal practice tools and models, legal training materials, a directory of expertise and other materials. Content is organized by practice area and content type and is searchable.

Access to external published research and data sources for evidence gathering is available through a third party legal research tool called LexisNexis Quicklaw, which gives lawyers access to a comprehensive collection of primary and secondary legal research materials, court decisions, legislation, legal commentaries, and current and archived news

IV-4: Evidence Gathering

While iCase has previously served as the government standard for assembling case documentation for evidence gathering, it is to be replaced by Microsoft's CRM Dynamic. The legal service unit of the Canadian Food Inspection Agency is already using CRM Dynamic successfully for this purpose.

JUS IT representatives indicated that there is a need to identify a faster content search engine for use by the Department; they are investigating the adoption of the Fast Search capability incorporated into Microsoft SharePoint 2013 as a possible solution. This would permit enterprise-wide indexing and search of JUS content and documents in any other repositories to which they have been granted access. Fast Search could potentially be used to create a cross-government Big Data search capability extending beyond JUS itself. During content processing, information can be written to a link database for subsequent use by an analytical capability in the software to calculate link popularity statistics and to perform relevance weighting of documents found. This could make relevant content more quickly available to JUS lawyers, improving their ability to assemble evidence to support their cases.

IV-5: Business Analytics

JUS has a Business Analytics group that uses SAS (Statistical Analysis System), a software suite developed by the SAS Institute that is used for advanced analytics, business intelligence, data management and predictive analytics. SAS can be used to retrieve and modify data from a variety of sources for the purpose of performing statistical analysis.

SAS Analytics is the main tool that is used to analyze data inputs from the various resource management tools in use in JUS, including IFMS, Peoplesoft, iCase, and other sources. Toundjer, Erman, the Director Business Management Strategic Planning and Business Management, believes that while JUS systems that provide operational information and statistics are functional, different systems produce different results. The current focus is therefore on fixing the data before moving forward with plans to enhance the systems to generate more meaningful data. The existing iCase timekeeping system is used for performance measurement.

Problems with the current environment that need resolutions as a precursor to implementing a Big Data approach in the business analytics area include:

1) There are some significant gaps in the current information:
   - An intake system is required to measure the demand for services;
   - The litigation system is not treated as a process, and therefore it is difficult to determine who is adding value;
   - There are 160,000 files on iCase, but a number of these are duplicate entries, or are initiatives that do not represent actual legal cases.

2) Non-chargeable hours aren't tracked, such as the provision of advisory services to clients, so the analysis is incomplete.

3) It is difficult to develop Key Performance indicators because of differences between reports and inconsistencies in the data that is reported.

4) Reliable data isn't available from the private sector for comparison regarding efficiency and performance of the department;

5) There is some internal resistance to providing the necessary data.

IV-6: Big Data Analysis

The JUS IT Department is investigating the use of various tools to perform Big Data analysis – such as HP's Autonomy which allows analysis of large scale unstructured Big Data repositories, and ROSS, an experimental artificial intelligence system built on IBM's "Watson" artificial intelligence platform developed by researchers at the University of Toronto. Although both systems hold promise for the future, they are still at very early stages in their development; any practical implementation of the tool is unlikely to occur for some time to come.

Section V: General and Future Trends

V-1: Future Trends: Possible Big Data Applications in Justice

Research has traditionally involved two fundamental steps - developing an initial hypothesis and finding proof that confirms or refutes the hypothesis. While this approach remains an appropriate research methodology, a new approach has emerged in the world of Big Data. Using artificial intelligence, massive stores of data can be searched for areas of correlation without using an underlying hypothesis previously identified by researchers. As an example, researchers used Google's intelligent search engines to identify a correlation between queries in its Google Trends web site and seasonal outbreaks of influenza in various countries (Google, n.d.).

Similar correlations are beginning to be discovered in the legal and judicial environments. For example, the correlation among outcomes of legal cases, judgments and appeals are beginning to provide the capability to predict the outcome of future cases. Also the correlation between massive stores of case evidence searched in electronic form by eDiscovery tools will provide key findings and evidence trends for use by legal counsel in trials.

Kevin Quinn, a former Assistant Professor of Government at Harvard, ran a contest comparing his statistical model to the qualitative judgments of 87 law professors to see which could best predict the outcome of all the US Supreme Court cases in a year. The law professors knew the jurisprudence and what each of the justices had decided in previous cases. They also knew the case law and all the arguments. Quinn and his collaborator, Andrew Martin collected six crude variables assembled from previous cases and analyzed the outcomes, which exceeded the lawyers' predictions. They concluded that whenever sufficient information can be quantified, modern statistical methods will outperform an individual or small group of people. (Shaw, 2014)

V-2: Predictive Analytics and Early Case Assessment

Lawyers make many strategic decisions and predictions during any stage of a trial based on their assessment of the outcome. Lawyers may also decide before taking a case to trial whether to negotiate a settlement offer. The ability to accurately predict the outcome of a case has practical consequences because litigation is risky, time consuming, and expensive. Errors in judgment can be costly in terms of time and resources, and also place a significant burden on the judicial system.

Many large legal firms are adopting the use of early case assessment tools and methodologies to estimate the risk of prosecuting or defending a legal case based on the financial costs and

resources required. Electronic legal discovery is also becoming increasingly costly. Organizations that spend significant resources on a case may eliminate the cost benefit of going to trial. Some organizations are also using the volume of information that can be produced to make cases more difficult and costly for the other side of the case to prosecute or defend.

Some existing software tools that can assist in and help facilitate the process of early case assessment include eDiscovery tools such as Exterro and Open Text eDiscovery. A US-based software company has also developed an application called "Picture It Settled", another example of a software tool used for early case assessment. This tool apparently uses neural networks, probability theory and behavioural patterns to predict the actions of opponents in a case, which can help to streamline negotiations. The software also estimates when parties are likely to settle and for what amount, with high accuracy. This doesn't replace legal judgement, but helps to understand alternatives and guide decisions by quickly modeling anticipated reactions.

Effective early case assessment requires a combination of professional expertise and software. Different resources in an organization typically use the software to assist in analyzing both structured and unstructured information[2] stored in electronic form. Depending on the sophistication of a case, lawyers may be assisted by IT professionals, forensic teams, and independent consultants. The tools used and the results of an early case assessment review can vary. Early case assessment is not a "one size fits all", but rather a process that needs to be managed and customized for each case.

The use of Big Data for case settlement and alternative resolution is expected to be one of the most significant future uses of Big Data in the judicial system. Information produced by the Data Analytics group in JUS indicated that the majority of cases processed by JUS are relatively small; in fact large cases are the outliers in statistical terms. While some cases processed by JUS must be taken to trial, many small cases may go to trial where the outcome can be predicted in advance. Significant savings in settling those cases without having to go to trial might result.

While JUS might be obliged to take a case to trial on principle, regardless of the possible outcome, predictive analytics may offer the opportunity to avoid trial in many situations.

[2] Structured data is organized in a highly mechanized and manageable fashion which can be easily processed by a computer, such as stored in Excel spreadsheets; by comparison, unstructured data, such as text found in e-mails and text reports is raw and unorganized. Searching through unstructured data can be expensive and difficult.

V-3: Current Limitations with Prediction Models

There are currently limitations to the predictive analytics approach to case outcome prediction and/or settlement. Predicting the outcome of new legal cases is still an imperfect science because of limitations of the current information is available for inclusion. e.g.:

- Cases may be settled without going to trial and aren't available for inclusion in the database, making the data incomplete;
- Courts may not have decided enough similar cases to permit the statistical prediction of case outcomes or feature weights that are needed to resolve the problem of small or biased samples;
- Algorithms that rely solely on assigning quantitative feature weights can be problematic because they are not sensitive to the particular context of a problem;
- The statistical algorithms used in the prediction models require sufficiently large data sets and, the more difficult the task, the more cases are needed to achieve accuracy;
- Text cases need to be represented in an appropriate form to enable machine learning; this is currently a largely manual process.

These difficulties are likely to be overcome with time and, given an appropriate database of cases, statistical or symbolic machine learning[3] techniques will be used effectively to determine general rules for classifying new cases and predicting their outcomes.

One major impediment to predictive analytics faced by JUS and the Canadian legal profession is the expense of building a complete and accurate Big Data store of cases and precedents. The information must also be kept current for new legal decisions and appeal results. It is unlikely that such a project could be funded in the near future without the backing of a consortium of law firms, or a third party organization such as LexisNexis who might make the information available by subscription. However, a detailed cost-benefit analysis would need to be performed before embarking on such a large project.

As a comparison, new regulations governing the accounting profession in 1999 forced the large accounting firms in the US and internationally to commission the development of a database containing information of all public and private companies, for use in determining possible conflicts of interest impairing auditor independence. Collectively, the firms engaged Sentinel, an organization supporting brokerage firms, to augment and modify their existing database of public and private organizations, and associated systems tools to accomplish this objective.

[3] Symbolic Machine Learning is another term used for predictive analytics or modeling where patterns of data are identified using human readable terms and symbols as opposed to numbers.

V-4: Data Management and Analytics

Controlling the information that is captured in large datasets can be problematic and subject to legal or ethical restrictions including:

- Documents used as evidence that contain personal information;
- Third party sources of information, such as articles or agreements that may be subject to copyright laws preventing open disclosure or dissemination;
- Confidentiality agreements where open disclosure could cause harm to a third party;
- Content compliance with government policies and practices.

Content management and curation of a JUS Big Data site will be an onerous task. Data will need to be kept current, as well as in compliance with laws and policies. Fortunately, software tools are being developed to assist with this process in the form of Data Management Solutions for Analytics (DMSAs). Gartner Group describes a DMSA as "a complete software system that supports and manages data in one or many disparate file management systems (most commonly a database or multiple databases) that can perform relational processing (even if the data is not stored in a relational structure) and support access and data availability from independent analytic tools and interfaces" (Gartner Inc., 2015). Organizations offering these tools include traditional IT firms, such as Teradata, Oracle, IBM, Microsoft, SAP and HP. However, new organizations, such as Cloudera, MapR, Actian and Pivotal are competing with the leaders.

Toundjer Erman, indicated that his objective was the "integration of information from all JUS systems that generate Enterprise Resource Management information in order to get a holistic view of all JUS operations." In parallel, there is a need to consider what the new operational landscape should look like, and then to generate new ideas by "looking through different lenses" and gaining new insights. This would include taking into consideration what other governments and public sector organizations are doing to use Big Data and technology to improve legal service processes and efficiency.

Ultimately, existing JUS data analytics information could be combined with other data for use in predicting how the legal environment will change. For example, will new legislation trigger more litigation, and what resources will need to be recruited or developed in JUS over a period of 3-5 years to respond to those predicted needs. Predictive Data Analytics can contribute to this analysis, but will require redevelopment of the current data architecture in the administration and resource planning areas to be more process driven.

V-5: Policy Development

Big Data offers an opportunity to contribute to government policy debates. Tools such as "Social Harvest" can extract data from Twitter, Facebook, and other social media platforms and log this information to a variety of data stores. Statistics Canada and many other government agencies possess a wide range of data concerning the behaviour of Canadians as a direct result of citizen interactions with government online services. However, a government department that uses social media to try to identify and better understand the needs of Canadians might also be accused of spying on its citizens in order to supress potential resistance.

The use of Big Data for policy development raises new moral and ethical issues for policy makers. Using predictive analytics and probability theory to predict what the general population might do in the future, as opposed to what they have done in the past could contribute to the policy debate. However, results based on findings from a relatively small group of people might still contain errors. A risk is that Big Data predictions about individuals might punish people for their propensities, not their actions, thus potentially denying basic human rights. Predictive analytics used by police in the US has led to a reduction in certain crimes, but resulted in the targeting certain socio-economic or cultural groups (Joh, 2014).

Section VI: Other Government Big Data Sources and Uses

Big Data offers a wealth of opportunities for other government agencies. Table 2 (below) shows a few of the areas in which Big Data is being exploited by other governments around the world.

Table 2: International Governmental Uses of Big Data

	US	UK	Australia	New Zealand	Singapore	Israel	Other
Predictive Policing	✓	✓	✓	✓	✓	✓	Spain
Informed Sentencing	✓	✓		✓			
Bail/Parole	✓	✓	✓			✓	
Fraud Detection	✓	✓			✓	✓	Canada
Health Care	✓	✓			✓	✓	Taiwan
Education	✓	✓		✓		✓	Korea, Canada
Public Works	✓				✓		Ireland, Philippines
Transportation	✓		✓	✓	✓		Sweden, Ireland
Infrastructure				✓	✓		
Economic Policy			✓				Japan, Germany, Canada
Environment				✓			Netherlands, Canada
Public Relations					✓		Japan, Hong Kong, China
Information Sharing	✓	✓			✓		Spain, Ireland, Japan
Government Resource Allocation			✓	✓	✓		Philippines, Germany

VI-1: Sentencing and Parole

Big Data can have a big impact on Correctional Services and on the Criminal Justice system in general by informing sentencing and parole decisions in a variety of ways.

Data-centered, evidence based strategies can be used to divert as many people as possible toward alternative programs, either within or outside of prisons, possibly reducing prison crowding and lowering the likelihood of re-offense. The Attorney General of the United States says "[d]ata can [...] help design paths for federal inmates to lower these risk assessments, and earn their way towards a reduced sentence, based on participation in programs that research shows can dramatically improve the odds of successful re-entry. Such evidence-based strategies show promise in allowing us to more effectively reduce recidivism" (Leopold, 2014). Similar risk assessments can be used to inform bail and parole decisions.

These types of strategies are being used effectively in a variety of jurisdictions:

The State of Florida and the province of Quebec both use statistical programs to profile juvenile offenders and assign them to risk-specific rehabilitation programs. These programs have shown significant success in reducing recidivism (Perry, McInnis, Price, Smith, & Hollywood, 2013).

The US states of Pennsylvania and Tennessee and the Australian state of New South Wales require statistical analysis to be used in all sentencing decisions;

The cities of Baltimore, Philadelphia and Washington, DC, all use algorithms to predict the likelihood of re-offence by parolees, and plan parolee supervision accordingly.

Big Data can also be used at a higher level to inform overall sentencing guidelines; the US Sentencing Commission is currently studying the use of data-driven analysis to issue general (not individual) policy recommendations. These could include changes in recommended sentence length where historical data shows current measures to be ineffective.

VI-2: Policing and Security

Law enforcement agencies have a history of using profiling and data mining to identify potential threats and predict criminal activity: Big Data offers a variety of tools to augment this capacity.

DEPLOYMENT

Predictive analytics are being used in over sixty major cities across the United States to help law enforcement agencies predict areas of probable criminal activity, and to assign patrols accordingly. These programs take into account times and locations of previous crimes, incident records, weather patterns, and historical and sociological information to create maps of "hot

spots". Cities using these maps have reported decreases of between 10% and 40% in criminal activity as a result. Los Angeles also tweets daily "hot spots" to citizens, to increase vigilance.

CRIME PREDICTION

Predictive analytics can also be applied more narrowly, to identify individuals at high risk of committing crimes. Chicago has a program in effect that uses a "heat list", created by a complex algorithm using data from a wide variety of sources. Officers or letters are sent to the homes of people on this list, to offer social services such as job training, or tailored warnings of increased penalties for certain crimes for people with particular prior convictions. The program has yielded positive results and is considered a success.

The U.S. Department of Homeland Security (DHS) and the Israel Security Agency (ISA) both have programs under development to detect terrorist attacks before they happen. DHS uses a Future Attribute Screening Technology to screen people for behavioural attributes associated with violent acts. Their Predictive Screening Project defines observable behaviours that precede a suicide bombing attack, and has shown promise in the testing phase. The ISA is investing in technology to convert unstructured data such as video and audio into a form that can be analyzed and used to produce real time alerts.

CRIME DETECTION

A third area of use for Big Data in policing is detecting crimes in near real time. This is being applied mainly to various forms of fraud, such as Medicare, securities, and bank fraud in the U.S. It is also being used in the UK to detect the misuse of prescriptions, and foreign bribery.

VI-3: Full Litigation Services

In 2013, Singapore launched a country-wide integrated Electronic Litigation System (iELS) for all litigation. iELS is accessible from anywhere through an internet browser, and has the following key functionalities (Braddell Brothers, 2015):

- Streamlining and re-engineering of high volume litigation processes;
- Information-based filing - Data capture (e.g. via XML and electronic forms) instead of only paper capture (e.g. document scanning), enabling the flexible re-employment of information as and when required;
- Active case management - Courts can pro-actively track and manage pending matters
- Litigation process management - Alerts and triggers designated to ensure that litigants do not miss critical deadlines;
- Electronic case file for lawyers - Lawyers have access to all relevant documents at any time and any place with an Internet connection, for the duration of each case;

- Integrated due diligence checks - Due diligence checks integrated with the electronic filing process, doing away with the need for subsequent back-room reconciliation;
- Court calendaring - Optimal assignation of court hearing days to be achieved with the syndication of date/scheduling information captured via information-based filing.

VI-4: Transportation

Intra and inter-city transportation systems (including both infrastructure and services) produce a vast amount of data from sources such as road sensors, bus GPSs, and ticketing systems that can be analyzed and used to increase the efficiency of services and allocate government resources. Some examples of foreign governments using these data to great advantage include:

- Swedish National Road Administration uses IBM systems to predict, control and optimize road traffic to improve air quality and reduce congestion. This resulted in peak-time road traffic congestion being dramatically reduced, air pollutants cut by up to 12 percent, and public transport usage increase significantly;
- The city of LA uses demand-responsive pricing for parking. Prices are based on data from parking sensors, surveys, weather forecasts, information about holidays, local business activities, etc.;
- The city of Dublin provides live road sensor and city bus GPS data to citizens, who can use it to plan their routes;
- Similarly, New Zealand uses predictive analytics to provide motorists with real-time information on traffic patterns via Variable Message Signs, in operation on highways across the country. These signs also display messages about accidents and road closures and conditions.

VI-5: Health Care

A large variety of health related data exists (patient records, genome information, successful/unsuccessful trials, hospital records etc.). By combining this data for analysis, variants of a disease can be identified, as well as subsets of patients who would benefit from different treatment plans. Following up with these groups could lead to better outcomes for the patients, and greatly advance the research, although this can be difficult if information is anonymized or de-identified. (President's Council of Advisors on Science and Technology, 2014)

Some examples of Big Data currently being used in the health care field include:

- New Jersey uses medical billing data to map out hot spots where there are the most complex and costly healthcare cases, as part of a program to lower healthcare costs
- The UK Food Standards Agency uses Twitter data to predict outbreaks in real time (often weeks before other methods);

- In Singapore, hospitals are using predictive analytics to predict relapses;
- Taipei Medical University analyzes and monitors performance across all hospitals.

VI-6: Economics

Reliable information about the current state of the economy is extremely important in making monetary policy decisions. Big Data can provide this information by predicting a wide variety of econometrics. For example, there are a variety of leading and lagging indicators of overall unemployment in a jurisdiction, such as automobile downgrades and decreased grocery spending (leading), and increased foreclosures and vacation cancellations (lagging). This sort of analysis can be used for early warning, real time awareness, and real time feedback for public policies and programs. (Letouze, 2012)

The Bank of Canada has suggested using existing monthly indicators in combination with big data to predict GDP growth before official quarterly National Accounts data are released providing more timely and accurate metrics to inform monetary policy decisions (Armah, 2013).

VI-7: Education

With the advent and increasing popularity of online learning, there are new sets of data available about how and what students learn, including responses to various new techniques and modes of delivery. Research into these data could yield great benefits to the field of education, including identifying what skills taught at which points in childhood, leading to better adult performance in certain tasks. Learning management systems (for use in actual classrooms) are also becoming more popular, and are adding to the available data. (President's Council of Advisors on Science and Technology, 2014)

Student data can also be used to identify and respond to student having educational difficulty. In 2012, Ontario's Ministry of Education identified 14,000 students across the province who had left high school with three or less credits needed to graduate. One year later, after a campaign to get them to go to summer school or take extra credit courses, 8000 of them had graduated (Solomon, 2013).

Section VII: Big Data and Privacy in Government

VII-1: Privacy Laws

There is a complex matrix of laws, regulations and practices that arise from the possible use of Big Data in Government, and might affect its usage. The major international, national, and provincial laws are summarized in Appendix XI-3~~XI-3~~. However, many other sector specific privacy laws and considerations exist that may also come into play, depending on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the type of consent obtained from the data subject, etc. The following observations can be made about the legislation:

Formatted: Hidden

- There is no single law or practice governing data privacy; legislation exists at all levels of government creating a complex matrix of international, national and provincial laws that govern the use of personal information in Canada and abroad.
- Although similar in concept, privacy laws are not always aligned; some laws also have cross-border and extraterritorial reach.
- Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA.
- Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may be in conflict with the Privacy laws.
- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes, and may therefore be difficult to apply.
- There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.

VII-2: Recent and Pending Changes to Privacy Legislation

All governments are struggling with ways to keep their data privacy legislation current, relevant, and usable in light of the rapid technological developments. Of particular concern are the new analytical tools that have the ability to mine data and analyze the ever-increasing data sources, and especially those that target personal information. Perhaps of even greater concern is the trend toward consolidation of existing databases into Big Data sources. The concentration of personal information from various sources adds complexity and risk. Privacy laws in the

international community are far from static, and changes are likely to have an impact on Canadian laws and practices as these occur.

"As business systems and processes become increasingly complex and sophisticated, organizations are collecting growing amounts of personal information. As a result, personal information is vulnerable to a variety of risks, including loss, misuse, unauthorized access and unauthorized disclosure. Those vulnerabilities raise concerns for organizations, governments and the public in general." (AICPA/CICA).

Recent changes made to legislation could have significant implications for personal data privacy and the rights of Canadians. The specific aspects of these laws are presented in Appendix XI-~~4~~XI-4. The laws include:

Formatted: Hidden

- Bill S-4 The digital Privacy Act
- Bill C-13 Protecting Canadians from Online Crime Act
- Bill C-51 Investigative Powers for the 21st Century Act (aka the "Anti-Terrorism Act")

VII-3: Legislative Restrictions, Guidelines and Safeguards Regarding Government Use of Personal Information

An increasing amount of information is available from the Canadian Government through its "Open Government" and other initiatives; this trend is likely to continue. At the same time, controls have been established to try to ensure that personal information is only made available to those who are authorized to access it.

ACCESS TO INFORMATION AND PRIVACY PROGRAM (ATIP)

Systems are controlled and information is subject to review under the requirements of the Access to Information Act and the Privacy Act before being released. The ATIP program also permits citizens to determine what information government holds about them, and provides them with the ability to correct the information if it is inaccurate.

Government procedures also exist surrounding the handling of personal information by its departments and agencies. Guidelines issued by the Treasury Board include a Directive requiring the performance of an extensive Privacy Impact Assessment (PIA) before implementing or changing government systems, or altering the manner in which they process information. The PIA includes guidelines for the assessment of privacy implications before entering into contracts or making outsourcing decisions.

THE STATISTICS ACT

The Statistics Act permits Statistics Canada to enter into contractual agreements to share confidential information with other government departments under specific conditions:

1) Information can be shared with the statistical agencies of provinces and territories for statistical purposes if:
    a. The data subjects were notified at the time of data collection;
    b. The provincial agency has the authority to collect the information on its own; and
    c. The agency's confidentiality protection requirements are substantially the same as those of Statistics Canada.
2) Where information is collected jointly by Statistics Canada and any federal and provincial government department, municipal government or other incorporated body such as an association or university, and where data subjects are notified in advance of intention to share the data, and are given the opportunity at the time of data collection to refuse to allow their information to be shared.

The OPC has also highlighted the existence of other laws that supplement, but do not necessarily supersede, the Privacy Act and PIPEDA and which provide Canadians with additional protections for their personal information:

"Several federal and provincial sector-specific laws include provisions dealing with the protection of personal information. The federal *Bank Act*, for example, contains provisions regulating the use and disclosure of personal financial information by federally regulated financial institutions.

Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals." Source (Office of the Privacy Commissioner of Canada)

Therefore, many substantial controls do exist over the internal use of personal information by government departments and agencies.

VII-4: Implications for the Department of Justice Canada

Determining which jurisdiction governs personal information is becoming much more complicated as information is gathered and/or transferred across legal jurisdictions and co-mingled in Big Data stores or linked with other information sources. It is also easy to lose track of the origin of the data over time, and especially if the organization operates across Canada or captures information on the Internet. Maintaining data accuracy and responding to citizen's

information requests becomes problematic. Courts around the world are struggling with data ownership and the determination of which laws will apply.

Most of the Big Data that is to be used by JUS is likely to consist of legal precedents and opinions, or possibly large quantities of evidence submitted in a court case to be analyzed using eDiscovery tools. Therefore, although there may be a few exceptions, (e.g. criminal records), JUS appears unlikely to capture and use a significant amount of personal Big Data, other than where it uses personal information contained in other government databases (e.g. StatsCan) for analytical purposes, and usually in aggregated form. However, the department may use personal information of its own staff members to assess efficiency and productivity of the various department functions. PIPEDA may also apply to aspects of litigation proceedings, depending on the context, when personal information captured in connection with litigation involves commercial organizations or is carried out in the course of commercial activities.

Regardless, JUS is likely to be involved in legal actions or discussions surrounding the use of personal data by other Government departments, and some of the evidence that it collects which includes sensitive or other personal information must be kept private. In these cases, JUS lawyers will need to respect their obligations under PIPEDA by ensuring that any personal information collected, used or disclosed in connection with any anticipated or actual litigation (or any other use) needs to be done either with the consent of the individuals, or must otherwise meet one of the applicable exceptions to the knowledge and consent principles of PIPEDA or the Privacy Act.

Section VIII: Privacy Concerns - Big Data and Government

The issues raised by the establishment of Big Data sources are not necessarily new, but relate to the difficulty of managing and protecting such large banks of information. Also, the sheer volume of the data held by government – both collectively and about each individual – creates the concern that profiles of individual characteristics and behaviour can be established that are quite complete and accurate. The application of predictive analytics to that information could permit the prediction of future trends and behaviours of both societies and individuals. While this might have benefits, there is a darker side to the existence of mass stores of personal data if the capability was misused.

The main concerns, discussed further below, are likely to be in four broad areas:

- Big Data Management and Security;
- Individual Consent regarding permitted uses of the personal information;
- Transparency and government disclosure of how data is collected, stored and used; and
- Lack of trust in government.

VIII-1: Big Data Management and Security

Large electronic sources of personal information can have significant value to those with less honourable intents. Once accessed, huge amounts of information can be rapidly transferred and stored inexpensively and with relative ease, attracting theft for monetary gain or extortion where personal exposure might have adverse impacts for both individuals and governments. The more attractive the information, the greater the difficulty to protect against data breaches by sophisticated hacking communities or tools – both in state-sponsored or private hands.

The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. This could involve greater risk of misuse in the event of a data breach, and eventual misuse for identity theft or fraud. The risk to government and individuals must be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

The demonstrated ability of hackers to overcome the security of government websites (e.g. recent attacks by Anonymous on Canadian Government web sites) and the perception that personal information is at risk of being disclosed or used fraudulently undermines public confidence in the safety of having their personal information in government data repositories.

"Each of the Canadian Privacy Statutes contains safeguarding provisions designed to protect personal information. In essence, these provisions require organizations to take reasonable technical, physical and administrative measures to protect personal information against loss or

theft, unauthorized access, disclosure, copying, use, modification or destruction. These laws do not generally mandate specific technical requirements for the safeguarding of personal information" (Piper, 2015).

Somewhat surprisingly, there are no prescribed standards for implementing security controls to protect personal information; rather it is left up to organizations to use their own judgement to determine what is appropriate. PIPEDA and the B.C. and Alberta privacy acts only "require organizations to take reasonable steps to safeguard the personal information in their custody or control from such risks as unauthorized access, collection, use, disclosure, copying, modification, disposal or destruction" (Office of the Privacy Commissioner of Canada, n.d.).

Reasonable safeguards include several layers of security, including, but not limited to risk management; security policies; human resources, physical and technical security; and business continuity management. The reasonableness of security arrangements adopted by an organization must be evaluated in light of a risk assessment including a number of factors, such as the sensitivity of the personal information; the foreseeable risks; the likelihood of damage occurring and the resulting harm caused; the medium and format of the storage method, and the cost of putting preventative measures in place.

VIII-2: Consent to Use Personal Information

The so-called "secondary use" of personal data - i.e. the use of data that has been provided for one purpose for other purposes - is a growing problem in the digital world, and in the Big Data world in particular. There is also a grey area between what information might require explicit or implicit consent for its use. The rules surrounding the requirement for consent and the use of personal information is clearly laid out for the private sector in PIPEDA, but Big Data will create broader issues for the public sector as well.

In the past, some of this data was considered to have been provided with the individual's implicit consent that it would be used in accordance with disclosures made by organizations. However, legislation covering the collection of most personal data collected by private organizations in Canada now requires explicit consent for use in accordance with specific terms. Any proposed secondary use for other purposes isn't generally permitted unless the use is disclosed at the time of collections. This is especially true in situations regarding the use of one of the sensitive categories of information (see Appendix XI-6~~XI-6~~).

Formatted: Hidden

Subject to legal interpretation, The *Privacy Act* might provide the government with more flexibility in its use of information provided to its various departments in the normal course of business, including the sharing and exchange of this information between government departments in the form of a Big Data repository, so long as the information is adequately protected from improper access or uses. Such use is already being made for research purposes

(e.g. by StatsCan). Sections 7 & 8 of the *Privacy Act* appear to cover this use. However, in the future expansion of Open Data and Big Data, where information is spreading out in many directions, it might be more difficult to determine whether information is being used in ways that don't require some form of additional consent or opt-out capability, and there may be unintended consequences. The standard form of consent or notification provided by the government will probably have to be worded very carefully at the front end of the process, and the back end of the process will require some form of careful review to ensure that the information is not being used outside of legal boundaries.

VIII-3: Transparency and Government Disclosure

The 2014 OPC survey reported, "The vast majority (89%) of those who had heard something about government surveillance activities agreed that surveillance or intelligence gathering agencies should have to explain their activities to Canadians" (Phoenix Strategic Projections Inc, 2014).

In 2000, the Canadian Government began to create its first Big Data repository, which became known as "Big Brother". The database included information on the addresses, education, marital status and ethnic origin of Canadians. It also tracked a person's employment and social assistance history, and their income tax records. Plans to implement the database were shelved at the time due to concerns expressed by the OPC and in Parliament, and also because of the volume of public requests to see their personal information contained in the database (CBC News, 2000).

The concerns of the Canadian public in this area remain today. A conclusion of the 2014 OPC survey was that "The majority of Canadians are not confidant that they have enough information to know how new technologies might affect their personal privacy." This would likely extend to the enhanced use of Big Data by government. "Canadians expressed varying levels of comfort with different ways in which government departments and agencies, including intelligence gathering organizations, could collect or share their personal information" (Phoenix Strategic Projections Inc, 2014).

Only about half of the OPC survey respondents felt that:

- They had a good understanding of what the government did with personal information that it collects;
- They were confident that the government would take their concerns about handling of their data seriously;
- They were confident that personal information shared with government would not be misused, lost or stolen.

VIII-4: Data Breaches and Trust in Government

While there are no statistics regarding trust in the Canadian government to protect personal information, numerous highly publicized data breaches have occurred in Canada over the past few years, and the numbers have grown substantially:

"The federal government reported breaching the privacy of individuals more than 5,000 times last year — an all-time high, according to new figures. The data are only for six departments, so the 5,237 privacy breaches they reported in 2014 are likely just a glimpse at what happened across government. Even so, the figure is almost as many as had been reported in the previous 11-year period, including instances where a taxpayer's or organization's information was incorrectly released, lost or compromised" (Press, 2015). Public awareness of attacks on government has increased too with the recent highly publicized attacks on Government web sites by "hacktivist" groups, such as Anonymous.

Canadians are waking up to the possible uses of their personal information by government agencies. The December 2014 OPC survey found that "56% of Canadians have some awareness of surveillance and intelligence gathering activities." "Roughly half (49%) of Canadians have seen, read, or heard something about surveillance or intelligence gathering activities for the purposes of national security in the past year or so" (Phoenix Strategic Projections Inc, 2014).

The heightened awareness of Canadians is likely a result of the recent publicity of government surveillance and information sharing programs through public revelations by Richard Snowden and the debate surrounding Bill C-51 (now the Anti-Terrorism Act) and its potential implications for the privacy of personal information. 78% of those polled in the OPC survey said they were either very (44%) or somewhat (34%) concerned about law enforcement and security agencies collecting their personal information for government surveillance purposes.

VIII-5: Big Data Privacy Controls

While the use of Big Data and related technologies can create significant privacy concerns as highlighted above, some of the technologies available now also permit the implementation of sophisticated controls to protect individual rights of citizens by regulating how Big Data technologies are used. Examples of these controls include:

- Use of methods for the "tagging" of data to ensure use is restricted to the purposes for which it was collected or generated;
- Implementing purpose-based or user-based controls according to the permissions and restrictions established for this data, including access controls;
- Tracking user access to data and the purposes for which it is used;

- Implementing algorithms that provide alerts regarding inappropriate access and possible uses.

While the use of specific information by JUS may not raise broad privacy concerns, other information available to government agencies may cause issues when data is aggregated or concentrated in electronic form, and especially when data is merged from multiple agencies. Regardless, there can be great benefits to merging and analyzing this information, such as:

- For research and public policy development regarding health, social, economic, national statistical trends;
- To demonstrate transparency and accountability of government; and
- To achieve public participation through engagement.

There is tremendous value in having broader access to this information for research, analysis, and policy development. Big Data is being used by the US Department of Justice to analyze medical billing records to detect Medicare fraud, and they are looking at similar Big Data sources for the detection of other frauds (Scannell, 2015). The increased sharing of information across government departments also creates complex relationships and can result in difficulties surrounding disclosure and transparency about the use of the information. One such example is the Canadian Open Government Portal that is intended to provide "greater transparency and accountability, increase citizen engagement, and drive innovation and economic opportunities through Open Data, Open Information, and Open Dialogue" (Government of Canada, n.d.).

Achieving full openness while maintaining appropriate controls over data privacy may be mutually exclusive objectives requiring some compromises. Legal privacy objectives can often be achieved through "de-identification" or "anonymization" of data, but the more heavily data is neutralized in this manner, the less useful it can become. In addition to legal requirements for compliance, there are also ethical considerations and, while privacy and confidentiality are somewhat different concepts, contractual and other agreements regarding the possible use of information (e.g. copyright) may need to be considered.

Focus groups during the 3rd International Open Data Conference held recently in Ottawa identified several privacy concerns and issues around open data:

- The public sector collects a great deal of sensitive personal information. While individual sources of anonymized or de-identified information might not reveal the identity of a person, the use of multiple data points that link or connect to others may make it possible to connect or triangulate between unrelated data points, making it possible to identify individuals.
- The use of Census and national statistical information can be problematic, even if data is aggregated, since individuals can often be identified within small groups or communities. Locational data can sometimes involve the same risk as a personal identifier "key", such as a name or social insurance number.

- The potential to profile, target or discriminate against vulnerable people or groups might be possible through matching of open data sources with information gained from other private sources.

The concern exists that while government surveillance will be made easier for protection against terrorism and illegal acts (Open Data Ottawa Privacy Conference Notes).

VIII-6: Forecasting Canadian Public Opinion on Privacy and Big Data in Government

According to a study conducted by the Office of the Privacy Commissioner (OPC) in December 2014, "Nine in ten Canadians expressed some level of concern about the protection of their privacy, with 34% saying they are extremely concerned (up from 25% in 2012)." Further, "Canadians increasingly feel that their ability to protect their personal information is diminishing. Seventy-three percent, the greatest proportion since tracking began, think they have less protection of their personal information in their daily lives than they did ten years ago" (Phoenix Strategic Projections Inc, 2014).

Canadians are therefore aware and concerned about the privacy of their personal information, and increasingly so. The primary focus of Canada's privacy programs has arguably been on the use of personal information in the private commercial sector, and the protection of this data through PIPEDA and its enforcement by the OPC. The same degree of knowledge or awareness of the Privacy Act and the permissions afforded by it to government doesn't seem to exist.

At the same time, recent legislative changes (see Appendix XI-4~~XI-4~~) and public revelations concerning clandestine government surveillance programs by Western governments, including Canada, have not likely helped to ease public concern. Some vocal members of the Canadian public, in particular, are questioning whether the extent to which the legislation is being implemented is commensurate with the need.

Formatted: Hidden

Anti-Terrorism Bill C-51, in particular, appears to be the subject of much concern. Daniel Therrien, the Privacy Commissioner of Canada, is responsible for the independent oversight of Canada's privacy laws and compliance. He recently submitted an article published in the Globe and Mail in which he said:

"In my view, Bill C-51, in its current form, would fail to provide Canadians with what they want and expect: legislation that protects both their safety and their privacy. As proposed, it does not strike the right balance.

The scale of information-sharing between government departments and agencies proposed in this bill is unprecedented. The new powers that would be created are excessive and the privacy safeguards proposed are seriously deficient" (Therrien, 2015)

The focus on the use of Big Data to track people and groups casts a negative image. The Commissioner's comments and position on information sharing are likely to create further debate and shape public opinion regarding government Big Data and data sharing between departments and with other governments. As sharing of Big Data information by government agencies becomes more commonplace, Canadians may become increasingly concerned about the possible uses, and react negatively.

Addressing the lack of awareness by Canadians of the way in which their personal information is being used may require greater emphasis on public disclosure to help reduce concerns. One recommendation made by the recent report to US President Obama suggests the implementation of a Consumer Privacy Bill of Rights based on the Fair Information Practice Principles. While this approach might help confidence in the private sector, a broader "Citizen's Bill of Rights" might be more appropriate to help renew the trust in government to protect personal information in the face of the expanded use of Big Data. One of the Key Informants, Howard Deane, from the Consumers Council of Canada, expressed the view the level of trust might be elevated if the government was more transparent regarding how personal information that makes its way into their hands will be used (i.e. limits on use), and what protections will be put into place around Big Data to avoid its misuse.

There are also ethical and moral questions about how Big Data might be used by government, or disclosed to others for possible misuse. There is a difference between government predicting and disclosing broad statistics about crime and cancer rates on a macro scale and using the data to focus in on individuals. The more granular the information becomes, the more organizations might be tempted to use the information in negative ways.

In his Globe and Mail article, the Privacy Commissioner indicated that the new legislation would "provide 17 federal government agencies with almost limitless powers to monitor and profile ordinary Canadians, with a view to identifying security threats among them. The end result is that national security agencies would potentially be aware of all interactions all Canadians have with their government. That would include, for example, a person's tax information and details about a person's business and vacation travel" (Therrien, 2015).

Public opinion is often difficult to predict because it often varies by culture, and is subject to "trigger" events that cause rapid shifts - e.g. The Edward Snowden disclosures surrounding government surveillance involved such a shift. Other than the OPC survey conducted by Phoenix Strategic Projections Inc., there appear to be few detailed Canadian surveys and public opinion polls that specifically address this topic in detail, but recent studies done on public

perceptions and opinions in the US and EU confirm that people believe that the privacy and security of their personal information is at risk, as is their ability to keep their information confidential in such an open world. However, there are a number of recent international studies that support this view.[4]

A Welcome Trust study in the UK found that focus group participants distinguished between acceptable types of government uses of personal data according to the following factors:

- The Government identifying needs, planning resources and services, and allocating funds;
- Prevention and detection of crime and, including terrorism;
- Identifying social/population trends and statistics;
- Unearthing dishonesty (e.g. fraudulent benefit claimants and tradesmen)

While there was a general awareness of data collection by both government agencies and companies generally, the Welcome study found that the public views of the collection and use of personal data could be summarized as follows:

- The public consider the collection and use of personal data to be a big issue;
- When asked, the public are ostensibly opposed to any form of data use and collection by government and companies;
- In practice, the public consider there to be no alternative to sharing personal information with government and companies in the modern world and expect this to increase in future;

A significant proportion of the public expected to feel less comfortable about sharing personal data in future.

VIII-7:Summary

The 2014 study commissioned by the President of the United States regarding Big Data and Privacy included the conclusion that:

"Although the use of Big Data technologies by the government raises profound issues of how government power should be regulated, Big Data technologies also hold within them solutions that can enhance accountability, privacy, and the rights of citizens." "Responsibly employed, Big Data could lead to an aggregate increase in actual protections for the civil liberties and civil

[4] - PEW Research Study - Public Perceptions of Privacy and Security in the Post-Snowden Era - November 2014
- White House Study - Big Data and Privacy Review - May 2014
- EU Byte Study - Report on public perceptions and social impacts relevant to Big Data - March 2014
- Eurobarometer Report - Attitudes on Data Protection and Electronic Identity in the EU - June 2011

rights afforded of citizens, as well as drive transformation improvements in the provision of public services" (Report to the Executive Office of the President).

It remains to be seen how the use of Big Data will translate into privacy concerns and the reaction by Canadians to the use of their personal information in Big Data repositories going forward. The level of trust in government, along with knowledge of why data is being collected and how it will be used also appear to be significant Issues, judging from recent public reaction to Bill C-51. There will likely be a need for programs to educate the public about these uses, and to promote the benefits, in order to establish a level of confidence and trust in the process, and to prevent a negative backlash such as occurred with the "Big Brother" database proposal in 2000.

Section IX: Major Findings and Conclusions

While Big Data is reforming many aspects of the world in which we live, the earliest successful models have been built on large databases of structured quantitative data, because this type of information is more easily and readily interpreted by binary computer logic. Although there are some exceptions, much of the information generated by the legal community or used in trials is unstructured data – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex and will take time to develop and refine.

Despite this, considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data elsewhere in the legal profession. The marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantages and cost efficiencies to those who adopt them.

Q1. Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

IX-1: Possible Big Data Strategy

JUS is in competition with other organizations in the legal community who will be making investments in these new technologies, and the Department will need to make similar investments, if only to be competitive and cost-effective as it conducts its business. It is involved in an "arms race", where all parties must move forward to avoid being placed at a strategic disadvantage. The strategic decision to be made by JUS is whether it should position itself as an early adopter of the technology, or be satisfied to be a "fast follower". The other decision will be how it should invest its limited resources to achieve its strategic objectives – i.e. what should the priorities be?

The one overarching conclusion that can be derived from the study is that large legal organizations that fail to plan for the implementation of these new technologies are likely to find themselves at a significant disadvantage from a competitive and cost-effectiveness standpoint. Donald Wochna, chief legal officer of Vestige Digital Investigations, was quoted in *Law Technology News* as saying: "Big Data in general, and predictive data analytics in particular, are the potential holy grail in the practice of law."

Q2. How can the Department of Justice adopt Big Data for its own needs?

IX-2: Possible Uses of Big Data and Predictive Analytics in JUS

The possible uses of Big Data by JUS, and the implications thereof, include:

Possible Applications of Big Data by JUS

The primary applications of Big Data analysis in the Department of Justice are expected to be the use of:

1) eDiscovery software tools to analyze and refine information in large databases of relevant documents for the production of evidence to be used in trials.
2) Predictive analytics and artificial intelligence to predict the outcome of cases based on a Big Data repository of precedents and legal opinions, which might ultimately be used to reduce time spent and effort devoted to settling cases or taking them through the trial process.
3) Data analytics techniques to analyze large databases of JUS operational statistics, with a view to improving individual performance and the overall productivity and cost-efficiency of the Department and, in future, to proactively position department resources to address emerging trends.
4) Data analytics to predict environmental trends, based on both internal (e.g. StatsCan) and external information (e.g. social media) that might be used to respond to the need for changes in government policies.
5) Automated tools for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process.
6) Automated tools to measure both individual performance and compliance with department and professional policies, procedures and standards and, in summary form, for management reporting of department performance and risk management.

Some Considerations Surrounding Big Data Implementation

Lawyers who have already been exposed to the use of eDiscovery, predictive analytics and other advanced technology tools are recognizing some of the implications as well as the potential opportunities of working with advanced technologies and applying these tools to large repositories of relevant data. However, this is still a relatively new concept for many in the legal profession, and so it is difficult for them to know where to begin with plans for implementation. The following issues will need to be considered:

7) The legal world is definitely headed down a path where sophisticated technologies (e.g. Big Data and Predictive Analytics) will play an increasing role. Legal organizations that

fail to keep up will eventually find themselves to be at a competitive disadvantage in terms of managing litigation cost and achieving success in the trial process.

The implementation and adoption of complex new technologies can be a significant undertaking in large organizations such as JUS, and therefore takes considerable time (i.e. years) to accomplish. Advance planning is therefore critical to ensure that resources are available, and the implementation is a success.

8) Implementation of the new technology tools and processes will require a strong change management program, based on the inherent resistance of people to significant change. The input we received, both in JUS and externally, is that such a program is likely to be required in order to achieve widespread adoption of new technologies, and also systems that attempt to measure individual performance more closely.
9) Significant investment will be required over many years, in money and human resources to remain current with the external legal marketplace to avoid falling behind. This is especially true with respect to the use of Big Data and predictive analytics technology where there have been, and will be, significant developments in the legal community.
10) JUS may not have access to the financial, human, and other resources required to move down all the emerging technology paths at once. The various options will need to be prioritized based on the projected cost/benefit before proceeding with any plans to implement Big Data, and considered as part of an overall departmental strategy.
11) Successful implementation is likely to depend on the ongoing commitment of JUS management to invest in the change, and to implement the tools required over a protracted period of time.

POSSIBLE COST EFFICIENCIES TO BE DERIVED FROM BIG DATA AND ADVANCED TECHNOLOGIES

The implementation of advanced technologies can be very expensive and disruptive to JUS, but the organization is likely to achieve both quality and cost-effectiveness improvements as a result. The following possible benefits were highlighted during our research:

12) Productivity and quality improvements would result from advanced expert search technology and litigation support tools to better research information and relevant evidence.
13) Possible process and efficiency improvements could be achieved in JUS administration and operations.
14) Productivity could be improved through the use of advanced analytics to allocate litigation resources by predicting forward demand and adjusting supply of legal resources accordingly.

15) Costs might be reduced through the ability to predict case outcomes and resolve cases through the use of an alternate dispute resolution mechanism involving the use of Big Data and predictive analytics to achieve a settlement without going to trial.
16) Access to internal and external Big Data sources would permit better policy decisions.

Is JUS Positioned to Take Advantage of New Technologies?

While the main purpose of this study was to look forward at possible uses of Big Data in JUS, the current uses of Information Technology in the Department was also reviewed. This is important as a starting point because the transition to the use of Big Data and predictive analytics in most large organizations relies on having a relatively strong base of technology on which to build. However, some technical and organizational restructuring may be required in order to move forward with more sophisticated technology programs.

JUS appears to have a variety of available technology tools, but there is some question as to how extensively these tools have been accepted and are being used by Department staff. In addition, some of the key systems (e.g. iCase) are aging and in the process of being replaced with Government standard tools, although implementation is just beginning.

Regardless, the conclusion is that:

17) No serious technology impediments were identified that would prevent JUS from moving forward with Big Data projects.

Q3. Are there promising practices in other countries and departments worth emulating? Where and what are they?

Practices in Other Countries and Departments

Sections D, E, and F of the report go into considerable detail about findings in this regard. The findings were mixed. Although there are some promising developments in other countries or departments that could be followed up, or developments to be followed, there do not seem to be any "magic bullets" at this time. However, there appears to be steady progress, and suppliers of technology in this area are making considerable investments.

18) Carrying on a "watching brief" while preparing to move forward as a clearer path emerges might be an appropriate strategy for JUS.

Q5. What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms,

incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms

IX-3: Government Big Data, Data Privacy and Public Opinion

PRIVACY LAWS

A complex matrix of laws, regulations and practices impact the possible use of Big Data in Government:

19) Canada is one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.
20) Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes.
21) There is no single law or practice governing data privacy; legislation that exists at all levels of government – a complex matrix of international, national and provincial laws exist that govern the use of personal information in the private and public sectors in Canada and abroad.
22) Although national laws are similar in concept, privacy laws are not always aligned; some also have cross-border and extraterritorial reach. Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.
23) Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may also be in conflict with the Privacy laws. Other laws and agreements must also be considered – e.g. copyright laws and contract laws may govern the use and disclosure of personal and other data.
24) Jurisdiction of data privacy laws and the determination of which applies depends on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the consent obtained from the data subject, etc.

Q4. What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

IX-4: Other Challenges to Privacy in a Big Data World

Data Security and Breaches

25. The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. Government programs to expand access to data through the Internet also create additional points of potential entry for breaches to occur.

While data contained in Big Data repositories is unlikely to be released in volume, the ability to access a wide range of views of various information sources through available portals, and potentially to use sophisticated search capabilities to retrieve information can be causes for concern if the appropriate level of security and controls aren't in place.

26. The risk to government and individuals will need to be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

Public Opinion and Reaction to Government Big Data

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become very concerned about their information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to sharing Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. There are a number of factors, in particular, that might trigger a negative public reaction or, conversely, steps might be taken to mitigate a negative reaction from occurring.

27. The implementation of a Big Data repository by government is likely to require greater government transparency about the way in which government handles personal information in Canada, and a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

Q6. What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

28. There will likely be a need to re-examine and revise the various laws affecting personal data privacy in Canada, and especially as government and other Big Data projects are brought on stream. In this regard, any changes to the legislation need to be forward thinking regarding emerging technologies (e.g. the laws need "to go where the puck is going to be" with privacy legislation, and not where the puck has been) otherwise laws will become quickly out-dated.

Individuals with legitimate access rights (e.g. government employees) who are able to download information can also be a source of concern if that information is lost or compromised. There are controls that can be put in place to partially guard against these sorts of occurrence, but they are generally expensive and cumbersome to implement.

Section X: References

AICPA/CICA. (n.d.). *http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/PRIVACY/GENERALLYACCEPTEDPRIVACYPRINCIPLES/Pages/default.aspx.*

Armah, N. A. (2013). *Big Data Analysis: The Next Frontier.* Bank of Canada.

Braddell Brothers. (2015). *Singapore Litigation Procedure.* Retrieved from Braddell Brothers: http://braddellbrothers.com/litigation.html

Brown&Ehrenreich. (2015, July 13). Can Big Data and Privacy Coexist?

CBC News. (2000). *Ottawa breaks up 'Big Brother' database.*

Dwoskin, E. (2014, August 22). Can Big Data Improve Medical Diagnoses? *Wall Street Journal.*

EDRM. (2015, 1 1). *www.edrm.net.* Retrieved 6 9, 2015, from EDRM.net: www.edrm.net

Gartner Group. (2014). *Magic Quadrant for E-discover Software.* Gartner Group.

Gartner Inc. (2015, June 14). *IT Glossary.* Retrieved from Gartner Group: http://www.gartner.com/it-glossary/big-data

Google. (n.d.). *Google flu trends.* Retrieved from goog.org flu trends: http://www.google.org/flutrends/

Government of Canada. (n.d.). Retrieved from Canadian Open Government Portal.

IBM. (2015, June 16). *What is Big Data.* Retrieved from Big Data at the Speed of Business: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Joh, E. E. (2014, February). *Policing By Numbers: Big Data and the Fourth Amendment.* Retrieved from Washington Law Review: SSRN: http://ssrn.com/abstract=2403028

Krasnyansky, A. (2015, January 29). *Meet Ross, the IBM Watson-Powered Lawyer.* Retrieved from PFSK Labs: http://www.psfk.com/2015/01/ross-ibm-watson-powered-lawyer-legal-research.html

Leopold, G. (2014). AG Says Big Data Can Reform Sentencing Rules. *HPC Wire.*

Letouze, E. (2012). *Big Data for Development: Challenges and Opportunities.* New York: UN Global Pulse.

Library and Archives Canada. (n.d.). Legislative Restrictions: Records of the Government of Canada.

Library of Parliament Research Publications. (2014). *Lgislative Summary of Bill S-4.* (L. o. Parliament, Producer) Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?ls=s4&Parl=41&Ses=2&source=library_prb&Language=E#a1

Library of Parliament Research Publications. (2015). *Legislative Summary of Bill C-51: Investigative Powers for the 21st Century Act.* Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?Language=E&ls=c51&Parl=40&Ses=3&source=library_prb

Microsoft acquires Equivio. (2015, January 20). Retrieved from blogs.microsoft.com: http://blogs.microsoft.com/blog/2015/01/20/microsoft-acquires-equivio-provider-machine-learning-powered-compliance-solutions/

Office of the Privacy Commissioner of Canada. (n.d.). *A Privacy Handbook for Lawyers: PIPEDA and Your Practice.* Government of Canada, Office of the Privacy Commissioner.

Office of the Privacy Commissioner of Canada. (n.d.). *https://www.priv.gc.ca/resource/fs-fi/02_05_d_15_e.asp.*

Office of the Privacy Commissioner of Canada. (n.d.). *Securing Personal Information: A Self-Assessment Tool for Organizations.*

Open Data Ottawa Privacy Conference Notes. (n.d.).

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.* Rand Corporation.

Phoenix Strategic Projections Inc. (2014). *2014 Survey of Canadians on Privacy.* Canadian Federal Government, Office of the Privacy Commissioner.

Piper, D. (2015). *Data Protection Laws of the World.*

President's Council of Advisors on Science and Technology. (2014). *Report to the President: Big Data and Privacy: a Technological Perspective.* Washington, DC: Executive Office of the President.

Press, J. (2015, March 22). Federal government privacy breaches soar to record high. *Ottawa Citizen.* Ottawa, Ontario, Canada.

Report to the Executive Office of the President. (n.d.). *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.*

Scannell, K. (2015, January 12). DoJ uses big data to crack Medicare fraud schemes. *FT.COM.*

Shaw, J. (2014, April). Why "Big Data" Is a Big Deal. *Harvard Magazine*.

Solomon, H. (2013, June 27). *How Ontario faces big data privacy challenges*. Retrieved from IT World Canada: http://www.itworldcanada.com/article/how-ontario-faces-big-data-privacy-challenges/47722

Therrien, D. P. (2015, March 21). *Without big changes, Bill C-51 means big data*. Retrieved July 2015, from Globe and Mail: http://www.theglobeandmail.com/globe-debate/without-big-changes-bill-c-51-means-big-data/article23320329/

Transparency Market Research. (2014). *eDiscovery Market Global Industry Analysis, Trends and Forecast 2014 - 2020*. Transparency Market Research.

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. University of St Andrews, UK.

Section XI: Appendices

XI-1: Current Industry Leaders in eDiscovery (Gartner Group, 2014)

- kCura markets the Relativity platform that supports collection, legal hold, processing, review, analysis and production of evidence. Relativity is sold through a wide range of service providers and hosting partners, and through a growing direct sales channel.
- FTI Technology, a separate business unit of FTI Consulting, offers both e-discovery software and services. Its main Ringtail platform performs functions from processing to evidence production. The Attenex product, also offered by FTI, provides a combination of machine learning and visual graphics for ease of document review.
- Recommind is known for its predictive coding technology, and supports all stages of the EDRM. Axcelerate eDiscovery can perform legal hold, collection, processing, review, analysis and production of documents, with Early Case Assessment and predictive coding capabilities.
- ZyLAB has an integrated solution supporting all stages of the EDRM. ZyLAB Intelligent Information Governance is used for file analysis and classification. Its e-discovery technology architecture is horizontally scalable and can handle large datasets.
- HP's Autonomy eDiscovery tool supports the full process of EDRM. Their self-service eDiscovery OnDemand model is part of an ongoing product development initiative that addresses the market shift toward organizations that want to bring e-discovery in-house. The product has a wide range of stakeholders ranging from IT users to in-house general counsel.
- Nuix's products include eDiscovery, Enterprise Collection Center, Web Review & Analytics, and Legal Hold. Its technology also extends to other related use cases, such as archive migrations, information governance and information security.
- Exterro provides products to support e-discovery from identification through review. Its primary offering is the Exterro Fusion E-Discovery software suite, which is built on a single open platform. Exterro's Fusion Integration Hub allows integration of existing legal, e-discovery and other information management systems.

XI-2: International Privacy Legislation

The original concept of data privacy was developed long before the explosion in the use of information technology could be envisioned. The impact of the new technologies used both in personal lives and in business is now apparent. The use of technology to access and manipulate personal data will continue, and is placing serious pressure on existing data privacy laws and practices around the world to keep up with the pace of change.

Political, geographical and cultural issues have made it difficult to adopt a single standard set of laws for data protection. Many different laws and regulations prescribe the privacy and treatment of personal information processed in Canada and in other legal jurisdictions. Most data privacy regimes include a range of seven to ten common principles. Those with a fewer number generally combine some of the principles, with a result that is largely the same.

The major forms of international legislation in place that prescribe the treatment of personal information from a data privacy standpoint are:

EU Privacy Directive

One of the original, and arguably the strongest of the international privacy regimes, is the EU Directive (Directive 95/46/EC of the European Parliament and Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data) enacted by the European Parliament in October 1995. The EU Directive forms the basis for most national data privacy regimes in place around the world today. Only countries with privacy regimes in place that are deemed adequate by the EU are permitted to receive personal information from EU countries. The Canadian public sector Privacy Act and private sector "Personal Information Protection and Electronic Documents Act" (PIPEDA) and their application have made Canada one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.

OECD Guidelines

The OECD Guidelines, issued in 1980 and revised in 2013, prescribe eight Basic Principles for National Application that align broadly with the EU Directive. The ten PIPEDA principles largely conform to the OECD standards and principles. The OECD principles follow:

1. Collection Limitation Principle

There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

2. Data Quality Principle
Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

3. Purpose Specification Principle
The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

4. Use Limitation Principle
Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except:

a) With the consent of the data subject; or

b) By the authority of law.

5. Security Safeguards Principle
Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.

6. Openness Principle
There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

7. Individual Participation Principle
An individual should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;

b) to have communicated to him, data relating to him:

i) Within a reasonable time;
ii) at a charge, if any, that is not excessive;
iii) in a reasonable manner; and
iv) in a form that is readily intelligible to him;

c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and

d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.

8. Accountability Principle
A data controller should be accountable for complying with measures which give effect to the principles stated above.

APEC Privacy Framework

The Asia Pacific Economic Cooperation (APEC) Privacy Framework provides for a flexible approach to information Privacy protection across member economies to avoid the creation of unrealistic barriers to information flows. The framework contains nine principles that are similar to the EU Directive, OECD Principles and PIPEDA. Privacy enforcement relies on authorities from participating APEC economies, including the Office of the Privacy Commissioner in Canada.

XI-3: Canada's Public and Private Sector Privacy Legislation

Canadian privacy legislation is aimed separately at the private and public sectors, and there are important distinctions between the two: the private sector includes privately owned, non-government entities, while the public sector includes organizations that are owned and operated by the federal, provincial, and municipal governments. Some examples are:

- Educational institutions such as universities, colleges, technical institutes, school boards;
- Provincial and regional health care institutions, nursing home operators, hospital boards and subsidiary health corporations;
- Local governments, including municipalities, police services and libraries.

Perhaps the most important difference between PIPEDA and the Privacy Act is that the former provides certain guarantees regarding the collection and use of personal information collected by private sector organizations, setting the stage for possible legal remedies and actions in the event of improper use and/or disclosure, whereas the Privacy Act does not set the same limitations on use of the information, nor does it provide for specific actions or remedies by government in the case of misuse, disclosure or data breach.

PUBLIC SECTOR PRIVACY LAWS - THE FEDERAL GOVERNMENT PRIVACY ACT

The Federal Privacy Act, first enacted on July 1,1983, applies to all of the personal information that the federal government collects, uses and may disclose about individuals or federal employees, i.e. it sets out policy surrounding the Government's collection, use and disclosure of their personal information in the course of providing services (e.g., passports, pensions, taxes).

The Privacy Act also sets out how federally regulated public bodies can collect, use, and disclose personal information, as well as how individuals can ask to access and update their personal information Examples of federally regulated public bodies include the:

- Bank of Canada
- Canada Revenue Agency (CRA)
- Canadian Space Agency
- National Research Council Canada
- Statistics Canada
- Treasury Board of Canada

The Act also gives the federal government a wide range of powers surrounding their possible uses and disclosure of personal information, subject to certain controls. The ability to share personal information across government agencies appears to be facilitated by the provisions of the Privacy Act. For example, Section 8 of the Act provides for disclosure in accordance with legal agreements between federal government departments, provinces and territories, First

Nations councils, and foreign governments for the purpose of administering or enforcing laws. Recent legislation further enhances the capability of sharing personal information across government agencies. See Appendix XI-4~~XI-4~~)

Formatted: Hidden

The Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with the Privacy Act.

PRIVATE SECTOR PRIVACY LAWS - THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC ACT

Federal legislation governing personal data privacy in Canada is provided in the Personal Information Protection and Electronic Documents Act (PIPEDA), which establishes the manner in which private sector organizations can collect, use or disclose personal information while conducting commercial activities in Canada. It also applies to the personal information of the employees of federally regulated organizations, such as telecommunications companies, banks and airlines. However, PIPEDA does not apply to non-commercial organizations such as charities or not-for-profits or political parties and some non-commercial associations.

PIPEDA generally applies to:

- Private sector organizations carrying on business in Canada in the provinces or territories, when the personal information they collect, use or disclose crosses provincial or national borders (except for the handling of employee information).

- Federally-regulated organizations with commercial operations in Canada, such as airlines, banks, telephone or broadcasting companies, but including their handling of health information and employee information.

PIPEDA sets out the following ten principles, which are closely aligned with the EU Directive, OECD Principles, and the principles adopted by the CICA and AICPA as "Generally Accepted Privacy Principles" (GAPP) (Appendix XI-5~~XI-5~~). The following privacy concepts are covered in the PIPEDA principles:

Formatted: Hidden

1. Accountability;
2. Identifying purposes;
3. Consent;
4. Limiting collection;
5. Limiting use, disclosure and retention;
6. Accuracy;
7. Security safeguards;
8. Openness;
9. Individual access; and
10. Compliance.

As with the Privacy Act, the Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with PIPEDA.

ROLE OF THE OFFICE OF THE PRIVACY COMMISSIONER OF CANADA

The Office of the Privacy Commissioner of Canada (OPC) advocates for the fundamental privacy rights of individuals through the establishment of an appropriate regulatory framework, and through the provision of independent oversight and monitoring of the application of PIPEDA to the private sector and the personal information handling practices of federal government departments and agencies to ensure compliance with the public sector Privacy Act.

OPC also acts as an ombudsman, working independently to:

- Advise individuals, government, businesses, and Parliament on emerging privacy issues;
- Investigate complaints and make recommendations based on findings; and
- Conduct audits under the two federal privacy laws; and
- Promote awareness and understanding of the protection of personal information.

PROVINCIAL LEGISLATION

Every province and territory has its own public sector legislation and these provincial acts also apply to provincial government agencies. Alberta, British Columbia and Québec have privacy legislation that is considered "substantially similar" to PIPEDA, so that the provincial act can be applied to private-sector businesses that collect, use and disclose personal information while doing business in those provinces. Ontario, New Brunswick, and Newfoundland and Labrador also have their own health care privacy legislation that supersedes PIPEDA in this area.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

Although provincial privacy laws are similar to federal laws, some important differences exist. For example, certain provincial privacy laws (e.g. Alberta) have special consent and transparency legislation that applies to organizations and/or their service providers who permit access to or disclose personal information to locations outside Canada. If this includes public sector bodies, it could create a conflict where information about an individual resident in a province might be shared with other governments, (e.g. in the case of suspected criminal, terrorist, or other activity, but where a crime has not yet taken place or been proven). It might also create problems with the potential capture, storage, and cross-border sharing of Big Data.

Only three provinces in Canada – Alberta, British Columbia, and Quebec - have their own private sector privacy legislation that supersedes PIPEDA; all others must comply with PIPEDA.

Organizations in Alberta, British Columbia, and Quebec therefore need to be careful of complying with both their own private sector privacy legislation as well as PIPEDA.

Alberta, Saskatchewan, Manitoba, Ontario, New Brunswick, Newfoundland and Labrador and Nova Scotia5have each passed health information protection laws to deal with the collection, use and disclosure of personal health information by public and private sector health care providers. Alberta and British Columbia have also passed privacy laws that apply to employee information. Some of these laws might not be considered to be sufficiently compliant with PIPEDA to be deemed substantially similar. Therefore, in some cases PIPEDA may still apply.

Some provincial sector-specific laws include provisions dealing with the protection of personal information. Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals.

PIPEDA doesn't apply to an organization where it operates entirely within a province that has privacy laws deemed substantially similar to PIPEDA, unless the personal information crosses provincial or national borders.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

(Office of the Privacy Commissioner of Canada)

XI-4: Recent Changes and Other Applicable Privacy Legislation

Bill S-4 The Digital Privacy Act

Bill S-4 amends the *Personal Information Protection and Electronic Documents Act*,2 the federal private sector privacy law. It does this in several notable ways, including by:

- Permitting the disclosure of an individual's personal information without their knowledge or consent in certain circumstances;
- Requiring organizations to take various measures in cases of data security breaches;
- Creating offences for failure to comply with obligations related to data security breaches; and
- Enabling the Privacy Commissioner, in certain circumstances, to enter into compliance agreements with organizations. (Library of Parliament Research Publications, 2014)

Bill C-13 Protecting Canadians from Online Crime Act

Bill C-13 deals with:

- The offence of non-consensual distribution of intimate images;
- Offences committed by means of telecommunication; and
- One aspect of the area of law, generally referred to as "lawful access", an investigative technique used by law enforcement agencies and national security agencies involving intercepting private communications and seizing information where authorized by law.

Bill C-51 Investigative Powers for the 21st Century Act (aka the "Anti-Terrorism Act")

Bill C-51 takes into account new communications technologies and equips law enforcement agencies with new investigative tools adapted to computer crimes. The new investigative powers within the legislation give law enforcement agencies the ability to address organized crime and terrorism activities online by:

- Enabling police to identify all network nodes and jurisdictions involved in the transmission of data and the ability to trace the communications back to a suspect. This includes information on the routing, but does not include the content of a private communication;
- Requires a telecommunications service provider to retain data to prevent its loss or deletion while law enforcement agencies obtain a search warrant or production order;
- Makes it illegal to possess a computer virus for the purposes of committing an offence of mischief; and

Enhances international cooperation to help in investigating and prosecuting crimes that extend beyond Canada's borders. (Library of Parliament Research Publications, 2015)

XI-5: AICPA/CICA Privacy Guidelines

The Ten Generally Accepted Privacy Principles

The ten Generally Accepted Privacy Principles are:

1. Management. The entity defines, documents, communicates and assigns accountability for its privacy policies and procedures.

2. Notice. The entity provides notice about its privacy policies and procedures and identifies the purposes for which personal information is collected, used, retained and disclosed.

3. Choice and consent. The entity describes the choices available to the individual and obtains implicit or explicit consent with respect to the collection, use and disclosure of personal information.

4. Collection. The entity collects personal information only for the purposes identified in the notice.

5. Use, retention and disposal. The entity limits the use of personal information to the purposes identified in the notice and for which the individual has provided implicit or explicit consent. The entity retains personal information for only as long as necessary to fulfill the stated purposes or as required by law or regulation and thereafter appropriately disposes of such information.

6. Access. The entity provides individuals with access to their personal information for review and update.

7. Disclosure to third parties. The entity discloses personal information to third parties only for the purposes identified in the notice and with the implicit or explicit consent of the individual.

8. Security for privacy. The entity protects personal information against unauthorized access (both physical and logical).

9. Quality. The entity maintains accurate, complete and relevant personal information for the purposes identified in the notice.

10. Monitoring and enforcement. The entity monitors compliance with its privacy policies and procedures and has procedures to address privacy-related complaints and disputes.

(AICPA/CICA)

XI-6: Other Categories of Personal Information

Sensitive Categories of Personal Information

Some personal information is considered sensitive. Sensitive personal information generally requires an extra level of protection and a higher duty of care. For example, some jurisdictions may require explicit consent rather than implicit consent for the collection and use of sensitive information. Some laws and regulations define the following to be sensitive personal information:

- Information on medical or health conditions
- Financial information
- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Sexual preferences
- Information related to offenses or criminal convictions

Source - (AICPA/CICA)

Nonpersonal Information

Some information about or related to people cannot be associated with specific individuals. Such information is referred to as *nonpersonal information*. This includes statistical or summarized personal information for which the identity of the individual is unknown or linkage to the individual has been removed. In such cases, the individual's identity cannot be determined from the information that remains, because the information is deidentified or anonymized. Nonpersonal information ordinarily is not subject to privacy protection because it cannot be linked to an individual. However, some organizations may still have obligations over nonpersonal information due to other regulations and agreements (for example, clinical research and market research).

XI-7: List of Key Informants

The following key informants were interviewed as part of the research process for this study

- Ms. Marj Akerley
  Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Kelli Brooks
  Principal in Charge, Evidence and Discovery Management
  KPMG LLP
  3020 Old Ranch Parkway, Seal Beach, California, USA
  USA

- Mr. Richard Cumbley
  Partner, Information Management and Data Protection
  Linklaters LLP
  1 Silk Street, London, United Kingdom

- Mr. Howard Deane
  Chair – Emerging Information Technology Committee
  Consumers Council of Canada
  1920 Yonge Street, Toronto, Canada

- Mr. Toundjer Erman
  Director, Business Management Strategic Planning and Business Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Dera J. Nevin
  Director of eDiscovery Services
  Proskauer
  Eleven Times Square, New York, New York, USA

- Mr. Chris Paskach
  Managing Director
  The Claro Group
  350 S. Grand Ave., Los Angeles, California, USA

- Mr. Jean-Sébastien Rochon
  Deputy Director and Counsel,
  National Litigation Support Services/National eDiscovery and Litigation Support Services

- Mr. Dominique Roy
  Director, Business Applications
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Julie V. Roy
  Supervising Counsel, National Litigation Support Services/National eDiscovery and Litigation Support Services.

- Ms. Tracy Sampson
  Deputy Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Dugald Topshee
  Director, Client Relationship Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Eric Ward
  Senior Counsel, Public Law Sector – Information law and Privacy Sector
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Dr. Anthony Wensley
  Associate Professor, Department of Management,
  University of Toronto Kaneff Centre, 3359 Mississauga RD N Mississauga, Canada

- Mr. Omid Yazdi
  Partner, Forensic Services
  KPMG LLP
  550 South Hope St. Los Angeles, California, USA

Tilhoff, Tanya

From: Topshee, Dugald
Sent: 2015-Oct-27 11:32 AM
To: * ISB-Management Team Justice
Subject: FW: Information on Big Data Workshop

Hi

Here are the details of the Big Data Workshop that I discussed at the management meeting today.

Dugald

From: Fraser, Charlotte
Sent: 2015-Oct-26 3:49 PM
To: Topshee, Dugald
Subject: Information on Big Data Workshop

Hi Dugald,

My Director, Alyson MacLean asked me to provide you with some details about the Big Data workshop to discuss at your management meeting tomorrow. We want to build on the research report that the contractors prepared by discussing the findings and options for Justice. We have secured a room at Library and Archives Canada for November 25th and are in the process of seeking approvals for hospitality. We have not come up with a full list of participants, and are seeking your input on who you think should be in attendance. Ideally, we would like everyone in ISB who met with our contractors earlier in the year to attend the workshop. We plan to contact Jeoffrey Bickert for some participants as well and will be inviting Toundjer Erman from Business Analytics. To ensure we have a meaningful discussion, we need a minimum of 15 Departmental representatives and a maximum of 25.

Below is the first part of the email message that we plan to send to participants once identified and I'm also attaching the draft agenda (currently being revised by our contractors).

Please let me know if you have any further questions.
Thank you
Charlotte

The Research and Statistics Division (RSD), Policy Sector has undertaken a research project on Big Data with E.S. Tunis and Associates in response to direction from the Deputy Minister to engage in forward-looking exercises on issues that may impact the Department (JUS) in the future. A research report was prepared and included the findings from interviews with both internal and external key informants. The findings indicate that large law firms that fail to plan for the implementation of new technologies will find themselves at a significant competitive and cost-effective disadvantage.

The Department needs to start thinking collectively about what JUS can do to prepare for the increasing use of Big Data technologies. To that end, the second phase of this project involves a one day workshop to consider the use of Big Data in the JUS environment and how JUS could respond to and advance some of the identified issues. The workshop is being held on November 25th from 8:30am to 5pm at Library and Archives Canada.

Big Data crosses all sectors of JUS, including litigation, IT, policy, business analytics, and planning. We need representatives from all relevant sectors to participate in a meaningful and collaborative discussion on the implications of the uses of Big Data in JUS.

Charlotte Fraser
Principal Researcher | Agente principale de recherche
Research and Statistics Division | Division de la recherche et de la statistique
http://canada.justice.gc.ca/eng/pi/rs/index.html | http://canada.justice.gc.ca/fra/pi/rs/index.html
Department of Justice Canada | Ministère de la Justice Canada
284 Wellington Street, Room 6115 |
284 rue Wellington, pièce 6115
chfraser@justice.gc.ca
Ottawa ON K1A 0H8
Telephone | Téléphone 613-948-3015
Facsimile | Télécopieur 613-941-1845
Teletypewriter | Téléimprimeur 613-992-4556
Government of Canada | Gouvernement du Canada

Justice Canada
Big Data Workshop

Wednesday, November 25, 2015

Library and Archives Canada

Objective: The objective of the workshop is to consider the research on forward trends and associated issues in the use of Big Data in the JUS legal environment and to consider what a Big Data Strategy for the department could be.

DRAFT AGENDA

9:00 to 9:10	**Welcome and Opening Remarks** ??? ??? will welcome participants to the Big Data Strategy Workshop, set the stage for the day and introduce the workshop facilitators.
9:10 to 9:30	**Introduction to the Agenda and the Strategic Planning Framework** Workshop facilitators from E.S. Tunis and Associates (ESTA) will review the strategic planning process and framework being used for this workshop.
9:30 to 11:30	**Big Data Research – What Issues does Big Data pose for JUS?** A research paper commissioned by the Research Division has examined the possible uses of Big Data in legal environments. It has identified some future trends and related issues for JUS to consider. Each set of trends and issues will be discussed and debated by participants.
11:30 to 12:30	**A Vision for Big Data in Justice Canada** Participants will be asked to construct a forward Vision for the positioning of Big Data in Justice Canada.
12:30-1:30	Lunch Break
1:30 to 2:30	**Building a Big Data Strategy** Participants will be introduced to a model of strategy building and will address the first set of issues from the Big Data Research
2:30 to 4:30	**Continuing** to build the Big Data Strategy Participants will break into small groups of their choice to articulate strategic responses to the remaining issues from the Big Data Research
4:30-5:15	**Big Data Strategy Review** Participants will present their strategies to the plenary session
5:15 to 5:30	**Next Steps** The facilitators will summarize the products of the day and outline next steps in the development of the Big Data Strategy for Justice Canada

Section 1: Activity Information (Mandatory) / Information sur les activités (obligatoire)

Activity Number / Numéro de l'activité	Core or Non-Core / Essentielles ou non-essentielles	What? / Quoi ? — Name or brief description / Nom ou description brève	Why? (explain why it is required) / Pourquoi ? (expliquer la nécessité) — Purpose (objective) of Activity / But (objectif) de l'activité	Rationale for # of Participants / Rationale pour le # de participants	Date — From / De	Date — To / À	Location / Emplacement — Venue, City / Lieu, ville
	Core	Big Data Workshop	workshop on developing a Big Data Strategy in response to direction from the DM to engage in forward-looking exercises on issues that may impact the Department in the future	all sectors of Justice, including litigation, IT, policy, business analytics, and planning and other government departments	25-11-2015	25-11-2015	Ottawa

Section 2: Travel / Voyage (Additional Travel Details / Détails supplémentaires de voyage)

# of Participants / Nombre de participants — Public Servant / Fonctionnaire	# of Participants / Nombre de participants — Non-Public Servant / Non-fonctionnaire	Name of Traveller / Nom du voyageur	Why not virtual presence? / Pourquoi ne pas utiliser la méthode virtuelle ?	Travel Category / Catégorie de voyage	Travel Type / Type de voyage	City / Ville — From / De	Mode of Transportation / Mode de Transport — Type / Catégorie	Mode of Transportation / Mode de Transport — Cost / Coût	Accommodations / Hébergement — Type / Catégorie	Accommodations / Hébergement — Cost / Coût	Meals / Repas — Cost / Coût	Incidentals and other costs / Frais accessoires et — Cost / Coût	Rationale for Transportation and Accommodations chosen / Justification pour transport et hébergement choisi	Estimated Expenditures / Estimation des dépenses — Total

Blanket Travel Authority and Annual Authority to Extend Hospitality / Autorisation générale de voyager et Autorisation pour des activités d'accueil annuelles (BTA and AAEH) — Yes / Oui No / Non	Estimated # of activities / # estimatif d'activités	Maximum Cost for all Activities / Montant maximal pour toutes les activités

Section 3: Hospitality / Accueil

# of Participants / Nombre de participants — Public Servant / Fonctionnaire	# of Participants / Nombre de participants — Non-Public Servant / Non-fonctionnaire	Host / Hôte — Name / Nom	Location / Emplacement — Facility / Installation	Location / Emplacement — Explain / Explication	Exceptional Hospitality Components / Éléments d'accueil exceptionnel	Food, Beverages and Other / Aliments, boissons et autres — Type / Catégorie	Forecasted Expenditures / Dépenses prévues — Cost per Individual / Coût par individu	Forecasted Expenditures / Dépenses prévues — Total Cost / Coût total
27	3	Justice Canada	Government facility / Installation gouvernementale			Other / Autres	$5 per break (2) and $23 for lunch = $33	$990.00

Section 5: Events - Other Expenditures / Événements - Autres dépenses and **Section 6: Other Comments / Autres commentair**

Type of Expenditure / Catégorie de dépense	Forecasted Expenditures / Dépenses prévues — Total Cost / Coût total	Description
meeting room rental, easels and equipment, consultants fee	$16,937.50	Hospitality below 1.5K (2 Refreshments and 1 Lunch)

Section 7: Summary / Sommaire

Total Forecasted Expenditures / Dépenses totales prévues	Funding / Financement — Expenses to be paid by / Dépenses à être payées par	Cost Centre / Centre de coût	Level of Approval Required / Niveau d'approbation requis	Approved or Declined/ Approuvé ou refusé (Initiales)	Contact / Ressource — Sector-Secteur	Contact / Ressource — Name / Nom	Contact / Ressource — Phone # / # de téléphone
$17,927.50	Justice (JUS)	70024	Level 2 / Niveau 2		Policy Sector	R-M Guerra	613-957-7949

Grand Total / Grand Total	$17,927.50	

Section 8: Approvals / Approbations

Expenditure Initiation Authority and Certified pursuant to Section 32 of the *Financial Administration Act* /
Pouvoir d`engagement des dépenses et Certifié en vertu de l'Article 32 de la *Loi sur la gestion des finances publiques*

FIRST NAME LAST NAME, TITLE, DIRECTION Initials Date

FIRST NAME LAST NAME, TITLE, DIRECTION Initials Date

FIRST NAME LAST NAME, TITLE, DIRECTION Initials Date

FIRST NAME LAST NAME, TITLE, DIRECTION Initials Date

Title of activity: Possible Big Data Uses by the Department of Justice and Related Privacy Concerns

Date(s), location: Existing Research Contract ending December 5, 2015

KEY ACTIVITY/MEETING/CONFERENCE (ACTIVITY)

Please describe the activity. (Is the activity one of policy or expenditure [i.e., FPT meeting, stakeholder consultation, conference, grant, contribution, or contract])?

This is a research activity. RSD has contracted with E.S. Tunis and Associates (ESTA) to investigate emerging trends affecting the current and possible future uses of Big Data and Privacy within the Department. This work is supporting direction from our Deputy Minister to engage in forward-looking exercises to explore issues that may impact the Department in the future. The first phase of the project consisted of research gathering – both from authoritative literature and interviews of Key Informants (including Departmental officials from IT, Public Law, Litigation Branch, and Business Analytics). ESTA provided RSD a draft report July 24th and colleagues in the above noted sections had an opportunity to comment on the draft.

The report notes the following possible uses of big data for Justice: for operational efficiency purposes (using case management and timekeeping systems), for policy purposes (using data analytics to predict environmental trends), and for litigation purposes (for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process). None of these potential uses would raise public concerns about individual privacy.

The second phase of the contract is to engage in a Departmental strategy session on possible uses of big data within the Department. The proposed plan is to hold a one day strategy development workshop with key Departmental officials and the contractors to consider the findings of this research and to explore what a big data strategy might look like for the Department. The intent is not to recommend the implementation of a Big Data Strategy, rather this workshop would consider the research evidence and what it means for Justice and how the Department could address or respond to issues surrounding the use of big data. The proposed timeframe for this Departmental event is the end of November.

What is the role of the officer who is assigned carriage of the file/attending the meeting/conference?

The role of the officers is to ensure the contractor submits their deliverables on time. Two officers will attend the departmental session/workshop.

Has there been or will there be a public announcement?
If yes, please describe below

[ ] Y [X] N

Y N

Title of activity: Possible Big Data Uses by the Department of Justice and Related Privacy Concerns
Date(s), location: Existing Research Contract ending December 5, 2015

Is the agenda attached? [ ] [X]

Who will be in attendance (if applicable, e.g., FPT officials, academics, public, NGO representatives)?

Participation will be from Justice officials from relevant sections including Information Solutions Branch, Public Law, Litigation Branch, and Business Analytics.

Justification for continuing the activity
Is it an activity that has one or more of the following criteria? (Check all that apply.)

- [X] *routine;*
- [X] *non-controversial;*
- [ ] *urgent and in the public interest;*
- [ ] *reversible by a new government without undue cost or disruption.*

Please describe the routine, non-controversial, urgent, etc., nature of the activity.

This activity does not involve any outside stakeholders except for the contractors.

If the activity does not fall within the above criteria, please explain.

Deferrable or Alterable
Can the activity be deferred or altered (e.g., amending an FPT agenda to remove sensitive issues, or deferring a contribution or a meeting/consultation)? Please explain:

Yes, the workshop/session can be delayed. The contract expires December 5th.

If the activity can be deferred or altered, please provide justification why the activity and/or the participation of the officer should continue:

Recommended approach

Provide the manager's recommendation for this activity.

It is recommended that the contract remain in place and that the internal Justice strategy development session/workshop take place in late November.

Title of activity: Possible Big Data Uses by the Department of Justice and Related Privacy Concerns
Date(s), location: Existing Research Contract ending December 5, 2015

Responsible manager:

Signature: _____

Date: 28 August 2015

Decision by SADM:

☑ Approved

☐ Not approved

☐ Approved subject to:

Signature: _____

Date: 28/08/15

Justice Canada
Big Data Workshop

Wednesday, November 25, 2015

Location: Library & Archives Building, Wellington Street

Objective: The objective of the workshop is to consider the research on forward trends and associated issues in the use of Big Data in the JUS legal environment and to consider what a Big Data Strategy for the department could be.

DRAFT AGENDA

8:30 to 8:40	**Welcome and Opening Remarks** JUS CIO??? ??? will welcome participants to the Big Data Strategy Workshop, set the stage for the day and introduce the workshop facilitators.
8:40 to 9:10	**Introduction to the Agenda and the Strategic Planning Framework** Workshop facilitators from E.S. Tunis and Associates (ESTA) will review the strategic planning process and framework being used for this workshop.
9:10 to 11:30 (With break at 10:00)	**Big Data Research – What Issues does Big Data pose for JUS?** A research paper commissioned by the Research Division has examined the possible uses of Big Data in legal environments. It has identified some future trends and related issues for JUS to consider. Each set of trends and issues will be discussed and debated by participants.
11:30 to 12:30	**A Vision for Big Data in Justice Canada** Participants will be asked to construct a forward Vision for the positioning of Big Data in Justice Canada.
12:00-12:30	Working Lunch
12:30 to 1:30	**Building a Big Data Strategy** Participants will be introduced to a model of strategy building and will address the first set of issues from the Big Data Research
1:30 to 3:45 (With break at 2:30)	**Continuing** to build the Big Data Strategy Participants will break into small groups of their choice to articulate strategic responses to the remaining issues from the Big Data Research
3:45-4:30	**Big Data Strategy Review** Participants will present their strategies to the plenary session
4:30 to 5:00	**Next Steps** The facilitators will summarize the products of the day and outline next steps in the development of the Big Data Strategy for Justice Canada

Annex C: Draft List of Participants to be invited to Big Data Workshop

	Name	Dept/section	Title	email	Phone number
1	Marj Akerley	JUS/Information Solutions Branch	CIO	**Marj.Akerley@justice.gc.ca**	(613) 941-3444
2	Dugald Topshee	JUS/Information Solutions Branch	Director, Client Relationship Management	**Dugald.Topshee@justice.gc.ca**	**(613) 952-8488**
3	Toundjer Erman	JUS/Strategic Planning and Business Management	Director, Business Management Strategic Planning and Business management	**terman@justice.gc.ca**	**(613) 948-5917**
4	Dominique Roy	JUS/Information Solutions Branch	Director, Business Applications	**droy@justice.gc.ca**	(613) 960-0876
5	Mala Khanna	JUS/Public Law Sector-Information Law and Privacy Section	Director and General Counsel	**mkhanna@justice.gc.ca**	**(613) 957-4624**
6	Eric Ward	JUS/Public Law Sector-Information Law and Privacy Section	Senior Counsel	Eric.Ward@justice.gc.ca	(613) 952-4130
7	Paul Vickery	JUS/Litigation Branch	Director General and Senior General Counsel	**pvickery@justice.gc.ca**	**(613) 948-1483**
8	Jean-Sébastien Rochon	JUS/Litigation Branch/National eDiscovery and Litigation Support Services	Deputy Director and Senior Counsel	**jean-sebastien.rochon@justice.gc.ca**	**(613) 670-6373**
9	Julie Roy	JUS/Litigation Branch/National eDiscovery and Litigation Support Services	Supervising Counsel	**julie.roy@justice.gc.ca**	**(613) 670-6359**

10	Stan Lipinski	JUS/Policy Sector/ Policy Integration and Coordination Section	Director General	**slipinsk@justice.gc.ca**	(613) 941-2267
11	Ryan Hum	PCO/Central Innovation Hub	Strategic Designer and Data Scientist	ryan.hum@pco-bcp.gc.ca	613-668-2193
12	Alan Bulley	ESDC/Strategic Policy and Research Branch	Senior Director	alan.bulley@hrsdc-rhdcc.gc.ca	819-654-1655
13	Peter Beaman	JUS/Legislative Services Branch	Deputy Chief Legislative Counsel (regulations)	peter.beaman@justice.gc.ca	613-957-0077
14	Gervais, Michèle	JUS/Change Management Office	General Counsel	michele.gervais@justice.gc.ca	613-946-6630
15	Katie Hammound	JUS/Information Solutions Branch	Director	**Katie.Hammoud@justice.gc.ca**	**(613) 941-5210**
14	Edward Livingstone	JUS/Public Law Sector	Director General	**elivings@justice.gc.ca**	**(613) 941-2326**
15	Yves Marion	JUS/Business Practices Division	Senior Director	**Yves.Marion@justice.gc.ca**	**(613) 957-4618**
16	Darlene Thibault	JUS/Management and CFO Sector	Director, Digital Workspace	**Darlene.Thibault@justice.gc.ca**	**(613) 952-5820**
17	Charlotte Fraser	JUS/Policy Sector/RSD	Principle Researcher	Charlotte.Fraser@justice.gc.ca	613-948-3015
18	Ting Li	JUS/ Policy Sector/RSD	Researcher	tli@justice.gc.ca	613-957-9584
23	Jill Wherrett	Public Safety Canada	Director General (Research, Planning, Policy)	Jill.Wherrett@ps-sp.gc.ca	
24	Lynn Barr-Telford	Statistics Canada	Director General, Health,	Lynn.Barr-Telford@statcan.gc.ca	613-951-1518

			Justice and Special Surveys		

Department of Justice Canada | Ministère de la Justice Canada

CCM#: 2015-012565
Classification: Protected B
FOR APPROVAL
Action by/Deadline: 2015/11/06

MEMORANDUM TO THE SENIOR ASSISTANT DEPUTY MINISTER

Travel Hospitality Conference Event Expenditure (THCEE) One-Off Request: Approval for the Big Data Workshop on November 25, 2015 (FOR APPROVAL)

SUMMARY

- A workshop will be held with Departmental officials on developing a Big Data Strategy. It will build upon the research report prepared as part of the body of work undertaken with the Deputy Minister.
- This work supports RSD's ongoing efforts to boost forward looking research capacity.
- The workshop will be held on November 25, 2015 in Ottawa. It is expected that 30 participants will be in attendance.
- The total cost of the event (including $990 for hospitality) is $17,927.50.
- It was approved by TAG on August 28, 2015 (Annex D).

DO YOU APPROVE under the Expenditure Initiation Authority and Certified pursuant to Section 32 of the *Financial Administration Act*? Please also sign the THCEE form under Section 8.

BACKGROUND OF THCEE ACTIVITY

The Research and Statistics Division (RSD) has undertaken a research project on Big Data with E.S. Tunis and Associates in response to direction from the Deputy Minister to engage in forward-looking exercises on issues that may impact the Department (JUS) in the future. A research report was prepared and included the findings from interviews with both internal and external key informants. The findings indicate that large law firms that fail to plan for the implementation of new technologies will find themselves at a significant competitive and cost-effective disadvantage.

The Department needs to start thinking collectively about what JUS can do to prepare for the increasing use of Big Data technologies. To that end, the second phase of this project involves a one day workshop to consider the use of Big Data in the JUS environment and how JUS could respond to and advance some of the identified issues.

Big Data crosses all sectors of JUS, including litigation, IT, policy, business analytics, and planning. Departmental representatives from all relevant sectors will be invited to participate in a meaningful and collaborative discussion on the implications of the uses of Big Data in JUS.

KEY CONSIDERATIONS / OPTIONS

A one day workshop will be held on November 25, 2015 in Room 156 at the Library and Archives in Ottawa developing a Big Data Strategy for the Department. Results from this workshop will be shared with participants as well as the Deputy Minister.

This is an opportunity for the Policy Sector to show leadership on a topic that impacts sectors across the Department.

A draft agenda is attached. Participants include representatives from the Information Solutions Branch, Litigation Branch, Business Analytics Unit, Public Law Sector, and Policy Sector.

FINANCIAL IMPLICATIONS

The estimated total hospitality cost is $990 [inclusive of taxes and gratuities] including a morning and afternoon break, and a working lunch.

DATE(S)	ITEM(S)	TOTAL COST
November 25, 2015	AM Break (30 x $5)	$150
November 25, 2015	Working Lunch (30 x $23)	$690
November 25, 2015	PM Break (30 x $5)	$150

The workshop was initially planned to be held over two days but with the budget restrictions in mind, the workshop was reduced to one day. Due to the full agenda, health breaks are being provided to avoid dispersing the group and lunch will be served as participants will be continuing their work through lunch. The per capita costs are below the average Treasury Board guidelines.

Other associated costs are as follows: meeting room rental ($402.50), flipcharts and equipment ($150) and the consultants (ESTA) ($16,385.00) who will plan, facilitate and prepare a final paper after the workshop. The total cost of the event (including hospitality) is $17,927.50.

The event respects all the rules, directives, and guidelines related to travel and hospitality.

COMMUNICATION PLAN

N/A

RECOMMENDATION

It is recommended that you indicate your concurrence by signing the approval block in the summary box.

Attachments
Annex A – Approval by the Transition Advisory Group
Annex B – Draft Agenda
Annex C – Draft List of Participants
Annex D – THCEE plan

Prepared by:
Rose-Marie Guerra, Event Planner and Coordinator, Intergovernmental and External Relations Division, 613-957-7949
Date: October 29, 2015

Reviewed by:
Charlotte Fraser, Principal Researcher, Research and Statistics Division, 613 948-3015
Date: October 29, 2015

Reviewed by:
Alyson MacLean, A/Director, Research and Statistics Division, 613-641-2266
Date: November 3, 2015

Reviewed by Direct Report to the ADM or ADAG:
Stan Lipinski, Director General, Policy Sector, 613-941-2267
Date: November 3, 2015

- ☐ The expense is within my budget.
- ☐ I confirm that funds will be committed.
- ☐ The policies and directives have been respected.

Department of Justice Canada | Ministère de la Justice Canada

CCM#: 2016 -001897
Unclassified
For Information

MEMORANDUM TO THE DEPUTY MINISTER

Outcomes of Big Data Research and Strategic Planning Workshop

(FOR INFORMATION)

SUMMARY

- The Research and Statistics Division (RSD) contracted with E.S. Tunis & Associates to conduct research on the applications of Big Data for the Department of Justice.
- This work was part of a collection of efforts to boost RSD's forward-looking research capacity and explore issues that may impact the Department in the future.
- This project involved two elements: (1) a literature review and interviews with key informants (internal to Justice and national/international experts); (2) a Big Data Strategic Planning Workshop with representatives from a variety of Sections/Sectors within the Department along with representation from Public Safety Canada, Canadian Centre for Justice Statistics, and the Privy Council Office.
- The report is attached at Annex A and the workshop outcome results are attached at Annex B.

BACKGROUND

In response to direction from the Deputy Minister to engage in forward-looking exercises on issues that may impact the Department in the future, the Research Statistics Division commissioned E.S. Tunis & Associates Inc. (ESTA) to conduct research into the applications and uses of Big Data within a legal context and what a Big Data strategy might look like for the Department. ESTA also explored the potential data privacy and protection implications associated with the use of Big Data by the Department.

The first part of the project involved a literature review of Big Data applications in Canada and abroad. This also included key-informant interviews with selected Departmental and Canadian/International experts on Big Data. A report was prepared and helped inform the discussion at the Strategic Planning Workshop hosted by RSD on November 25th, 2015 (Attached at Annex A).

The Big Data Workshop was attended by 20 officials, with representation from the Information Solutions Branch, Litigation Branch, Legislative Services Branch, Communications Branch, Information Law and Privacy Section, Finance and Planning

Branch, Business Analytics Unit as well as representation from the Canadian Centre for Justice Statistics, Public Safety Canada, and the Privy Council Office (a list of attendees is included in Annex B).

The research conducted by ESTA found that there are no serious technological impediments that would prevent the Department from moving forward with Big Data projects. The one overarching conclusion that can be derived from this work is that large legal organizations that fail to plan for the implementation of new technologies are likely to find themselves at a significant disadvantage from a competitive and cost-effectiveness standpoint.

DISCUSSION

The attached outcomes report highlights six strategies for the Department to advance opportunities for the use of Big Data:

- Improve Data/Information Transparency;
- Become a Big Data "Fast Follower";
- Effectively Resource Big Data in JUS;
- Protect Privacy;
- Improve the Ability to Access and Use Data in JUS; and
- Improve Data Sharing.

Advancing Big Data opportunities within the Department requires collaboration and cooperation between multiple sectors.

RESOURCE

(N/A)

COMMUNICATION IMPLICATIONS

(N/A)

NEXT STEPS

The Research and Statistics Division (RSD) is exploring options to develop a pilot project to advance some of the ideas discussed during the workshop. For example, RSD will consult Litigation Branch on the feasibility of analyzing data used in existing evidence management software (e.g., Ringtail) for predictive analysis.

Multiple sectors need to be engaged in order to advance Big Data in the Department. RSD will consult with the relevant Sectors to seek opportunities for collaboration in this area.

There are also opportunities to use Big Data in policy development. RSD will continue to work with Statistics Canada and Public Safety Canada to prioritize projects that would benefit from leveraging and combining multiple data sources. This is an item that could be raised when the Deputy Ministers of Justice and Statistics Canada meet in a couple of months.

Attachment: Annex A: Possible Big Data uses by the Department of Justice and Related Privacy Concerns
Attachment: Annex B: Big Data Strategic Planning Workshop Outcomes Report

Prepared by:
Ting Li, Researcher, RSD, Policy Sector, (613) 957-9584
Date: January 28, 2016

Reviewed by:
Kelly Morton-Bourgon, Principal Research, Policy Sector, (613) 957-9600
Date: January 29, 2016

Stan Lipinski, Director General, Policy Sector, (613) 941-2267
Date: January 29, 2016

Approved by:
Donald K. Piragoff, Senior Assistant Deputy Minister, Policy Sector

Possible Big Data Uses by the Department of Justice

And Related Privacy Concerns

September 1, 2015

Prepared By:

2B-268 FIRST AVENUE OTTAWA, ON CANADA K1S 2G8

T 613 594 3033, F 613 594 8928

info@estaconsulting.org

www.estaconsulting.org

Table of Contents

Section I: Executive Summary

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends". What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement.

Big Data is reforming many aspects of today's world, and organizations everywhere are finding ways to use it to achieve competitive advantage. The Canadian government will eventually be obliged to adopt Big Data applications in order to remain internationally competitive. An overall strategy, which considers all of the relevant issues, would help the Government of Canada and all of its departments and agencies to harness Big Data while fulfilling its responsibility to protect the public.

The Department of Justice (JUS) asked E.S. Tunis & Associates Inc. (ESTA) to conduct research into applications and uses of Big Data being made in legal and justice systems that might be considered for use by JUS, and what a Big Data strategy might look like for the department. ESTA was also requested to consider the potential data privacy and protection implications associated with the use of Big Data by the Department. The research method included a review of primary, and secondary sources both internal to JUS and external. Following is a summary of the research findings.

BIG DATA APPLICATIONS IN THE JUSTICE SYSTEM

Considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data in the legal profession. A large part of the information generated by the legal community or used in court proceedings is in electronic form, but much of this is unstructured – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex; they have taken time to develop and refine, but they are now coming rapidly on stream. In fact, the marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantage and cost efficiencies to those who adopt them.

There is widespread and growing use in western countries of intelligent eDiscovery software tools to analyze and refine large files of relevant documents for the production of evidence to be used in trials. New sophisticated analytics programs are emerging in the US to accurately predict case outcomes without the need to go to trial. Other potential uses of Big Data applications for consideration by JUS might include:

- Techniques to enhance and analyze large operational databases such as the JUS Case Management and Timekeeping systems to improve JUS productivity and cost-efficiency and, in future, to manage resources to address emerging trends.

- The use of data analytics to predict environmental trends using internal (e.g. StatsCan) and external (e.g. social media) information to contribute to policy debates.

- The use of automated tools and Big Data sources for early identification of risk associated with individual legal cases, and to manage risk throughout the trial process.

Promising innovation is also taking place in the justice systems of other countries. Singapore launched a countrywide Integrated Electronic Litigation System for all litigation to optimize scheduling of court dates, streamline court filings, and provide case management manage high volume litigation. iELS is accessible from anywhere through an internet browser.

The development and adoption of a Big Data strategy by JUS will not be a simple or inexpensive undertaking. It will take careful planning and a long-term commitment if the strategy is to be successful. It will not be possible for JUS to stand still in this area. JUS lawyers will find themselves at a competitive disadvantage to other lawyers in courtrooms, and these pressures will inevitably initiate change. The strategic decision to be made is whether JUS will be an early innovator or a "fast follower". Either way, a careful planning and budgeting exercise will need to be undertaken.

JUS already makes use of technology applications that will provide it with a strong base from which to move forward with the deployment of Big Data and predictive analytics systems. Over the long term, it is predicted that the implementation of Big Data applications will provide both quantitative and qualitative benefits to JUS.

DATA PRIVACY CONSIDERATIONS

Almost by definition, the concept of Big Data in government runs contrary to the concepts of personal data privacy, because a Government Big Data repository must ultimately contain a great deal of personal information about its citizens. Even if a data source is carefully screened to ensure that data is appropriately "de-identified", these protections may disappear when the data is combined with other sources for other uses., This raises potential privacy concerns and, depending on the situation, the potential for negative public opinion.

The implementation of Big Data systems by JUS specifically, and the Canadian government generally, will be challenging from a number of different standpoints.

- There is no single law or practice governing data privacy across Canada; different laws govern the privacy of personal information in the Public and Private Sectors and legislation exists at all levels of government. Although many laws are similar in concept, they are not always aligned, leading to a complex matrix of legislation and practices that surrounding the use of personal information in the private and public sectors in Canada.

- Some privacy laws also have cross-border and extraterritorial reach. Canada is one of the few countries accepted by the EU as having adequate data privacy protections for personal data transfers from the EU. While this status speaks to the strength of Canada's privacy laws, it needs to be preserved as it gives Canada an economic advantage over the many other trading nations that do not have the same status.

- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes. These laws may need change or, at a minimum, to be reconciled as to how they apply in practice.

PUBLIC OPINION ABOUT GOVERNMENT BIG DATA AND DATA PRIVACY

Canadian public opinion about government use of Big Data mainly surrounds how their information will be used and protected. Canadians will be concerned with the security and privacy of their information held by government:

- From an IT security standpoint
- From a transparency standpoint (having knowledge of what is being done with their data).
- From a trust in government standpoint.

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become concerned about their personal information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. The implementation of a Big Data repository by government is likely to require greater government transparency about the way in which

government handles personal information in Canada, and also a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

SUMMARY

There is probably no alternative to the future use of expanded Big Data applications and repositories by JUS. It will become an imperative, if the operation of the Department is to remain cost-effective and competitive. Ultimately, the privacy concerns that arise from the use of Big Data by JUS, on its own, are likely manageable. The implications of the increasing and much broader capture and use of Big Data by government in general creates a number of legal, policy and other issues that JUS will inevitably need to help to resolve as it moves forward with Big Data applications.

Section II: Introduction

II-1: The Evolution of Technology and Big Data

The explosion of computing, electronic sensing and digital communications technology in today's society has led to an exponential growth in online data; we create roughly 2.5 quintillion bytes of data a day, so much that an estimated 90% of the data currently in existence were created in the last two years (IBM, 2015). Large, growing subsets of this mountain of data are referred to as Big Data.

The term "Big Data" has a variety of definitions. For this study, we have defined it as "vast data sets that, when analyzed by algorithms, may reveal patterns, associations, and trends. In particular, these findings relate to human behavior and interactions. For the most part, these are datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Brown&Ehrenreich, 2015). What all sources agree on is that Big Data is defined by some combination of size, complexity, and technological requirement (Ward & Barker, 2013).

Big Data repositories are a result of the exponential increase in the amount of data being captured, combined with advances made in low cost digital storage media. Almost all transactions are now done online, and most documents and forms are now available in digital form only. Internet-enabled devices that are capable of capturing personal, environmental and geolocational data surround us. This data is being used and combined in increasingly innovative ways that were often not anticipated during the initial collection process. Governments and private organizations alike are beginning to recognize the value of this data, and are investing heavily to harvest it to gain a competitive edge and other strategic advantages.

As data repositories have expanded and evolved, so too have the methods and processes that permit data search and manipulation. The cost of storage has decreased to the point where much data is kept indefinitely, often because it is easier and cheaper to do so than to devote resources to culling it. In the meantime, advances in processing power and the creation of new ways to combine and analyze the data have permitted the combination and parsing of the data for novel uses.

This new flood of information has led to a large number of opportunities across a wide variety of sectors, while at the same time giving rise to some new privacy concerns as more data is gathered in a world where many electronic devices are now internet enabled, and are beginning to monitor and store information on almost everything that we do. Given enough data about an individual, it is possible to create a very detailed profile that removes all prospects of future privacy.

II-2: The Inherent Conflict between Big Data and Data Privacy

The data that exists in a Big Data world must ultimately include a great deal of information about real people. In the past, this information existed in siloes that were, for the most part, physically separated because information was stored in paper files. Even in the early days of electronic data processing, information was stored on devices that were only accessible by individual computers with no connection between them. In today's world of Big Data, these electronic files are capable of being linked both physically and logically together to permit broader information access and greater system functionality.

Clearly, there is growing value in harnessing Big Data. Predictive modelling using Big Data sources will permit doctors to make more accurate medical diagnoses (Dwoskin, 2014). Medical diagnostic programs may soon be capable of using Big Data findings to review a patient's entire medical history, X-Rays and results of medical tests online – from virtually any location in the world – to make a diagnosis. Vendors can use multiple sources of information to predict retail trends and match their supply of goods and resources with anticipated demand. Governments can monitor health and other emerging social trends in their countries to forecast the need for public programs, resource allocations and budgeting.

An individual's privacy has long been considered a fundamental human right. However, the Canadian Charter of Rights and Freedoms, when enacted in 1982, didn't anticipate a world where an individual's personal information could be captured and stored in such minute detail, nor the ways in which it might need to be specifically protected. Sections 7 and 8 of the Charter have often been interpreted to provide these protections, but may not provide the required degree of specificity in a world where the various permutations and combinations of the data make it very difficult to ensure individual anonymity.

One of the first big uses of analytics applied to Big Data sources in government has been by the intelligence community, which developed programs such as Carnivore[1] to monitor and analyze large amounts of electronic communications in order to detect subversive activities. Predictive analytics are also being used to forecast crime levels based on regional and local demographics. This information is also being used, primarily in the US, in predicting an offender's likelihood of reoffending as a basis for sentencing decisions. Big Data history is already being used to predict future population trends. As more data is captured about the everyday activities of individuals, it will not only be possible to make predictions about their health and welfare as a basis for improvement, but also whether they may be more susceptible to committing criminal acts before they commit them. The Big Brother world of George Orwell's "1984" might have arrived.

[1] Carnivore was a system implemented in the US in 1997 by the Federal Bureau of Investigations to monitor email and electronic communications sent over the Internet

While there are ways to disguise personal information in large data sources, the current focus of investment and research is mainly on ways to harvest Big Data, rather than on how to protect it, along with an individual's data privacy. An appropriate balance will need to be established if Big Data and privacy are to co-exist peacefully in Canadian society.

II-3: Why Big Data?

Regardless of the challenges, Big Data offers considerable opportunities to the Department of Justice (JUS). In order to take advantage of the opportunities, and to minimize the negative effects of Big Data, JUS needs to develop a clear picture of the current state and likely near-term evolution of the technology. To this end, JUS commissioned ESTA Consulting to conduct research into:

- The impact of Big Data in the context of current privacy laws in Canada;
- Ways in which JUS could adopt Big Data for its own needs;
- The implications/opportunities of Big Data including the possible role for JUS, and whether a "Big Data Strategy" would help; also more generally for the Government of Canada.

Implementing a Big Data strategy is not a simple task, especially for organizations the size of JUS or other federal public departments. All organizations now use Information Technology (IT) to a greater or lesser extent, but it is important for organizations to understand their current use of technology as a prerequisite for planning how they might move forward. The purpose of this report is therefore threefold:

1) To broadly review the current applications of IT by the Department of Justice in order to help it assess its position vis-à-vis other legal organizations with respect to the implementation of the advanced technologies and techniques employed by others to harness the power of Big Data in the public and private legal sectors;
2) To identify some of the privacy issues from a legal or regulatory standpoint that might stem from the availability and use of Big Data by JUS and the federal government, both currently and in the foreseeable future;
3) To consider the implementation of Big Data by the Government of Canada and some of the broader issues that could arise from this use, including public opinion and reaction.

Section III: Methodology

III-1: Project Scope

JUS requested the following scope in the form of questions to guide the direction of the research:

SECTION 1: HOW THE DEPARTMENT OF JUSTICE CAN MAXIMIZE THE USE OF BIG DATA?

Q1. Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

Q2. How can the Department of Justice adopt Big; Data for its own needs?

Q3. Are there promising practices in other countries and departments worth emulating? Where and what are they?

SECTION 2: PRIVACY

Q4: What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

Q5. What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms, incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms.

Q6. What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

III-2: Research

The following research activities were undertaken:

INITIAL RESEARCH

Initial research was performed to assist with the scope of the research, and in planning. This involved an initial review of material available online, and meetings and discussions with JUS research staff members to clarify roles and responsibilities. Initial research also included a literature review to identify existing and near-future uses of Big Data in the legal sector, both

private and governmental. The literature review also helped to better define the scope of the research.

PRIMARY RESEARCH

Primary research consisted of interviews with Key Informants, both internal and external, to identify issues and help establish an accurate overview of the industry, as well as to assist in determining the criteria to be used in analyzing the results and developing the conclusions. Selected Key Informants' views were solicited to help shape some research, highlight background issues and subject matter, and provide some assistance with key observations and conclusions. Their comments, where relevant and notable, were included verbatim in the report. (See Appendix XI-7 for the complete list of Key Informants interviewed).

Key Informants were also asked for their views and observations on the subject of potential Big Data uses in JUS and the Federal Government, and on the associated data privacy issues and public perception, in order to identify the issues and to establish a general overview of the environment and the potential issues and concerns. A Key Informants plan and guide was assembled to direct discussions with the informants, but questions were modified for each interview to match the particular area of expertise of the Key Informants. The focus of the questions was on the direction of Big Data development in the legal marketplace, and possible data privacy implications associated with the use of Big Data, both by JUS, and more broadly by the Canadian Government.

In order to gain an understanding of Canada's use of Big Data in relation to the rest of the world, research was also conducted into the uptake of the identified technologies in various jurisdictions. An overview of other ways foreign governments use Big Data was also established.

SECONDARY RESEARCH

A broad background and view of the environment, drivers, issues, and industry players was developed from the initial and primary research. Published reports, research papers, websites, Internet sources on the topics, together with media reports, were then examined. Further research was then conducted into each of the identified Big Data uses in order to understand their capabilities, limitations and methods of use, and to identify the most common product options in use. Secondary research focused on areas of legal administration related to the business of JUS.

All research was conducted with a view to produce an initial identification of emerging issues and risks in the use of Big Data, primarily by the JUS, but also more generally by government agencies.

Drafts and the final reports were reviewed with JUS staff to ensure accuracy and to verify scope coverage.

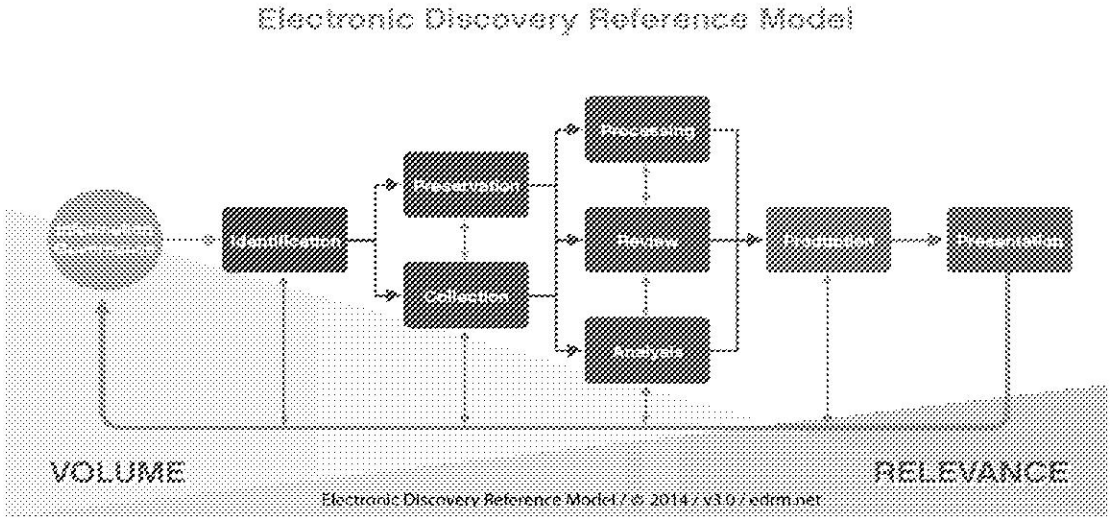
Section IV: The Emerging Uses of IT In the Field of Law

While technology applications, such as practice management systems (e.g. for time-keeping and financial management) have been used for some time in legal service organizations, the broader use of technology tools has been a relatively recent development. This has likely been driven in part by the explosion in the amount of unstructured (i.e. text-based) information in electronic form, and partly by innovations in the technology world to improve the ability to search, correlate and interpret this unstructured information in meaningful ways to gather and interpret evidence used in legal cases.

This section discusses some of the rapidly evolving uses of technology in the legal profession, including enhancements attributable to the emergence of Big Data; the sophisticated tools used to analyze large stores of data in the areas of eDiscovery and evidence gathering; legal research; the prediction of trial risk and outcomes and in the use of advanced data analytics for practice management and to achieve productivity improvements.

IV-1: eDiscovery Methodology and Tools

Electronic discovery (eDiscovery) tools include software designed and used to identify, preserve, collect, process, review, analyze and ultimately to produce information in electronic form to support the legal discovery process as legal cases are being conducted. E-discovery software capabilities include the ability to identify, preserve, collect, process, review and produce information for use by counsel. These capabilities are generally conducted in a sequential order prescribed in the Electronic Discovery Reference Model (EDRM, 2015), a framework that has been established and is broadly accepted by eDiscovery practitioners.



During the initial phases of the eDiscovery process, the data collection and early assessment capabilities of eDiscovery software is used to refine the data so that an initial evaluation can be made regarding the information quality, the location of information that is available for use in a case, and what additional resources might be required for its effective evaluation. A risk assessment is generally performed during this stage to determine whether any restrictions might govern the use of the data, such as policies or data protection laws.

Subsequent phases of the eDiscovery process generally include technology assisted review tools that employ analytics-based machine learning technology. These use statistical techniques to "train" the software to review the electronic files, thus reducing the required amount of manual review to improve overall cost-effectiveness of the review process.

eDiscovery tools have evolved considerably since they were first introduced to the legal marketplace. In its May 2015 "Magic Quadrant for eDiscovery Software" study, Gartner Group (Gartner Group, 2014) studied 18 of the top organizations providing eDiscovery solutions and services to the marketplace today. They positioned the 7 organizations described in Appendix XI-1 as the current industry leaders.

Key Informant Kelli Brooks, who heads up KPMG's Evidence and Discovery Management Group in the US, noted that the kCura Relativity platform is the most commonly used tool, but indicated that the following eDiscovery platforms had potential for creating significant developments in the eDiscovery industry:

- Equivio is a relatively new Israeli text analysis start-up company that was bought by Microsoft in 2015. Industry speculation is that Microsoft plans to integrate the Equivio machine learning technology into Office 365 in future.
- Brainspace is a revolutionary new tool that can be used to reveal complex relationships between documents for review.

PREDICTED CHANGES IN THE EDISCOVERY MARKETPLACE

Transparency Market Research, a U.S.-based provider of syndicated research, customized research, and consulting services estimated that the Global eDiscovery market was valued at USD 5.56 billion in 2013. Government and regulatory agencies were the largest end-user segment in 2013, accounting for about 51% revenue share of the global eDiscovery market. They expected the market to grow at a cumulative annual rate of 15.5% from 2014 to 2020 as eDiscovery solutions find widespread applications in government and regulatory agencies, small, mid and large-sized enterprises and law firms. (Transparency Market Research, 2014)

The eDiscovery marketplace will also change as electronic evidence expands from the current analysis of email, documents and voice mail to include social media and mobile data. Increases

in data transfers between inter-connected business systems will require growth in the ability to analyze structured data. A combination of human skill and sophisticated software tools such as predictive coding and structured data analytics will be required to analyze these more complex evidence streams. A number of large players in the IT world are investing heavily in both eDiscovery software and tools for predictive analytics in the legal marketplace. These include:

HP Autonomy – The Hewlett-Packard purchase of the Autonomy search engine in 2011 for $10.3 billion set the stage for their entry into the eDiscovery marketplace, and they have continued to invest heavily since in new functionality (e.g. a cloud-based offering) to expand in the legal marketplace;

ROSS – a result of collaboration between the University of Toronto and IBM, using IBM's Watson Artificial Intelligence engine for legal research (Krasnyansky, 2015);

Microsoft's purchase of the rapidly growing Equivio in January, 2015 for a rumoured $200 million gave them access to "a provider of machine learning technologies for eDiscovery and information governance. We are making this acquisition to help our customers tackle the legal and compliance challenges inherent in managing large quantities of email and documents." (Microsoft acquires Equivio, 2015).

Key informant Dera Nevin advised that two important issues must be addressed before an organization can move forward with plans to capitalize on the use of Big Data for eDiscovery or more sophisticated applications (e.g. Artificial Intelligence (AI) and Predictive Analytics):

1) An appropriate information governance structure must be in place so that the organization has knowledge of what electronic information they have and where it is stored. In the past, legal organizations have been overly reliant on the use of paper documents, and a significant cultural change is required to overcome this issue.
2) Organizations need to standardize on a limited set of eDiscovery tools to permit legal counsel to become experienced with their use. Lawyers won't become experts in programming, but they will need to become adept in future at using sophisticated tools to search for and manipulate data.
   Although software standardization is a desired goal, Ms. Nevin observed that large legal organizations like the Department of Justice are also exposed to a wide variety of legal scenarios, and since there are specific strengths and weaknesses of the various tools on the market, a single eDiscovery solution might not be suitable for every case. The need to differentiate between structured and unstructured data may also require different tools.

STATUS OF THE USE OF EDISCOVERY TOOLS IN JUS

JUS IT representatives indicated that the Ringtail tool, from FTI Technology, is used across the Government for evidence management. In addition to JUS, the RCMP, PCO, PPSC, Election Canada, and the Treasury Board apparently use Ringtail. However, at least one of the key informants we spoke to expressed the view that JUS use of the tools was not as extensive as it might be, and that JUS might be lagging the private sector in this area.

Jean-Sébastien Rochon and Julie Roy of the National eDiscovery and Litigation Support Services group indicated that about 14-16 paralegal positions are devoted to the support of Ringtail and eDiscovery tools. Ringtail is only used for files involving more than 5000 documents as it is not cost-effective on smaller cases.

A problem highlighted with the current implementation of Ringtail is that documents from the 1700 cases stored in the system are held in "silos", so documents used for evidence in one case aren't available for use in others, although they might be useful. Going forward, Rochon and Roy hope to restructure the Ringtail database so that over 25 million pages of documents could be searched across the system and made available if they are relevant to other cases, and aren't subject to legal privilege.

Another problem they identified was that legal units assigned to other government departments sometimes use other eDiscovery tools not recommended by JUS. These create files that aren't compatible with JUS and therefore can't be shared.

USE OF EDISCOVERY IN OTHER JURISDICTIONS

A review of other jurisdictions finds mixed approaches to the application of eDiscovery. Table 1 (below) shows an overview of the use of eDiscovery and governing laws in various countries:

Table 1: EDiscovery Around the World

Country	eDiscovery Legislation	eDiscovery Use
Canada	• Sedona Canada Principles Addressing Electronic Discovery (1st ed 2008, 2nd ed 2015) (Federal, compatible with all provinces and territories except Quebec, based on US) • Ontario, Nova Scotia, Manitoba, Saskatchewan, Alberta and BC all have guidelines for eDiscovery based on the Sedona principles • Quebec, as a civil law province, has different rules	• Widespread • Following the American example

United States	• Legislation in effect since 2006 (meet and confer), updated 2007, 2015 (pending)	• Widespread • Pioneering and exporting eDiscovery to the world
United Kingdom	• Legislation in effect since 2009 (meet and confer), updated 2013 • Very specific eDiscovery guidelines and requirements • Litigation budget is required early in the process	• Widespread • Some jurisdictions require all cases to use eDiscovery, some allow the judge to make the decision on a case by case basis
Australia	• Legislation in effect since 2009 (meet and confer) • Update in judge training program includes managing eDiscovery and electronic case management	• A court can order all discovery for a case be done electronically. • Most courts have implemented individual guidelines specifically for eDiscovery
New Zealand	• Legislation in effect since 2012 (meet and confer) • All discovery is now electronic, unless the court decides otherwise.	• Waited a long time to make rules, and had a chance to see what other commonwealth countries did • EDiscovery is now ubiquitous
Japan	• No laws governing eDiscovery for domestic litigation.	• Not very common in non-governmental cases • An expectation of data production exists for government investigations • Slowly gaining popularity, mostly driven by international litigation, particularly with US law firms and vendors
Korea	• No specific eDiscovery laws • Very strict privacy laws, including a requirement that all corporate and personal data be hosted domestically	• Virtually non-existent
Singapore	• Integrated Electronic Litigation System (iELS) implemented in 2013	• All cases use eDiscovery through the iELS

IV-2: Timekeeping, Document and Case Management

JUS has used its proprietary iCase tool for a number of years to store documents used in litigation. iCase is also used for time and case management. The JUS IT group indicated during our interview that a major goal is to align JUS systems to the extent possible with prescribed federal government standards. GC Docs has been adopted as the standard for record keeping and document management, with Microsoft SharePoint 2013 as the front-end interface and

system portal. JUS will be converting, but implementation is only in the early stages, with migration of existing content occurring in the later stages.

IV-3: Legal Research

Internal and external information sources are used for legal research. *Justipedia* is the central legal knowledge management repository for the Department of Justice. It contains legal opinions, pleadings and facta, agreements and other precedents and tools. It is also used to access legal practice tools and models, legal training materials, a directory of expertise and other materials. Content is organized by practice area and content type and is searchable.

Access to external published research and data sources for evidence gathering is available through a third party legal research tool called LexisNexis Quicklaw, which gives lawyers access to a comprehensive collection of primary and secondary legal research materials, court decisions, legislation, legal commentaries, and current and archived news.

The Canadian Legal Information Institute (CanLII), a non-profit organization managed by the Federation of Law Societies of Canada, also offers a free legal database that is rapidly becoming one of the research tools of choice. CanLII provides lawyers with access to court judgments, tribunal decisions, statutes and regulations from all Canadian jurisdictions.

IV-4: Evidence Gathering

While iCase has previously served as the government standard for assembling case documentation for evidence gathering, it is to be replaced by Microsoft's CRM Dynamic. The legal service unit of the Canadian Food Inspection Agency is already using CRM Dynamic successfully for this purpose.

JUS IT representatives indicated that there is a need to identify a faster content search engine for use by the Department; they are investigating the adoption of the Fast Search capability incorporated into Microsoft SharePoint 2013 as a possible solution. This would permit enterprise-wide indexing and search of JUS content and documents in any other repositories to which they have been granted access. Fast Search could potentially be used to create a cross-government Big Data search capability extending beyond JUS itself. During content processing, information can be written to a link database for subsequent use by an analytical capability in the software to calculate link popularity statistics and to perform relevance weighting of documents found. This could make relevant content more quickly available to JUS lawyers, improving their ability to assemble evidence to support their cases.

IV-5: Business Analytics

JUS has a Business Analytics group that uses SAS (Statistical Analysis System), a software suite developed by the SAS Institute that is used for advanced analytics, business intelligence, data management and predictive analytics. SAS can be used to retrieve and modify data from a variety of sources for the purpose of performing statistical analysis.

SAS Analytics is the main tool that is used to analyze data inputs from the various resource management tools in use in JUS, including IFMS, Peoplesoft, iCase, and other sources. Toundjer, Erman, the Director Business Management Strategic Planning and Business Management, believes that while JUS systems that provide operational information and statistics are functional, different systems produce different results. The current focus is therefore on fixing the data before moving forward with plans to enhance the systems to generate more meaningful data. The existing iCase timekeeping system is used for performance measurement.

Problems with the current environment that need resolutionas a precursor to implementing a Big Data approach in the business analytics area are:

1) There are some significant gaps in the current information:
   - An intake system is required to measure the demand for services;
   - The litigation system is not treated as a process, and therefore it is difficult to determine who is adding value;
   - There are 160,000 files on iCase, but a number of these are duplicate entries, or are initiatives that don't represent actual legal cases.

2) Non-chargeable hours aren't tracked, such as the provision of advisory services to clients, so the analysis is incomplete.

3) It is difficult to develop Key Performance indicators because of differences between reports and inconsistencies in the data that is reported.

4) Reliable data isn't available from the private sector for comparison regarding efficiency and performance of the department;

5) There is some internal resistance to providing the necessary data.

IV-6: Big Data Analysis

The JUS IT Department is investigating the use of various tools to perform Big Data analysis – such as HP's Autonomy which allows analysis of large scale unstructured Big Data repositories,

and ROSS, an experimental artificial intelligence system built on IBM's "Watson" artificial intelligence platform developed by researchers at the University of Toronto. Although both systems hold promise for the future, they are still at very early stages in their development; any practical implementation of the tool is unlikely to occur for some time to come.

Section V: General and Future Trends

V-1: Future Trends: Possible Big Data Applications in Justice

Research has traditionally involved two fundamental steps - developing an initial hypothesis and finding proof that confirms or refutes the hypothesis. While this approach remains an appropriate research methodology, a new approach has emerged in the world of Big Data. Using artificial intelligence, massive stores of data can be searched for areas of correlation without using an underlying hypothesis previously identified by researchers. As an example, researchers used Google's intelligent search engines to identify a correlation between queries in its Google Trends web site and seasonal outbreaks of influenza in various countries. (Google, n.d.).

Similar correlations are beginning to be discovered in the legal and judicial environments. For example, the correlation among outcomes of legal cases, judgments and appeals are beginning to provide the capability to predict the outcome of future cases. Also the correlation between massive stores of case evidence searched in electronic form by eDiscovery tools will provide key findings and evidence trends for use by legal counsel in trials.

Kevin Quinn, a former Assistant Professor of Government at Harvard, ran a contest comparing his statistical model to the qualitative judgments of 87 law professors to see which could best predict the outcome of all the US Supreme Court cases in a year. The law professors knew the jurisprudence and what each of the justices had decided in previous cases. They also knew the case law and all the arguments. Quinn and his collaborator, Andrew Martin collected six crude variables assembled from previous cases and analyzed the outcomes, which exceeded the lawyers' predictions. They concluded that whenever sufficient information can be quantified, modern statistical methods will outperform an individual or small group of people. (Shaw, 2014)

V-2: Predictive Analytics and Early Case Assessment

Lawyers make many strategic decisions and predictions during any stage of a trial based on their legal risk assessment of the strength of their legal position and the likelihood of a positive outcome. Lawyers may also decide before taking a case to trial whether to negotiate a settlement offer. The ability to accurately predict the outcome of a case has practical consequences because litigation is risky, time consuming, and expensive. Errors in judgment can be costly in terms of time and resources, and also place a significant burden on the justice system.

Many large legal firms are adopting the use of early case assessment tools and methodologies to estimate the legal risk of prosecuting or defending a case based on the financial costs and resource required. Electronic legal discovery is also becoming increasingly costly. Organizations that spend significant resources on a case may eliminate the cost benefit of going to trial. Some organizations are also using the volume of information that can be produced to make cases more difficult and costly for the other side of the case to prosecute or defend.

Some existing software tools that can assist in and help facilitate the process of early case assessment include eDiscovery tools such as Exterro and Open Text eDiscovery. A US-based software company has also developed an application called "Picture It Settled", another example of a software tool used for early case assessment. This tool apparently uses neural networks, probability theory and behavioural patterns to predict the actions of opponents in a case, which can help to streamline negotiations. The software also estimates when parties are likely to settle and for what amount, with high accuracy. This doesn't replace legal judgement, but helps to understand alternatives and guide decisions by quickly modeling anticipated reactions.

Effective early case assessment requires a combination of professional expertise and software. Different resources in an organization typically use the software to assist in analyzing both structured and unstructured information[2] stored in electronic form. Depending on the sophistication of a case, lawyers may be assisted by IT professionals, forensic teams, and independent consultants. The tools used and the results of an early case assessment review can vary. Early case assessment is not a "one size fits all", but rather a process that needs to be managed and customized for each case.

The use of Big Data for case settlement and dispute resolution processes is expected to be one of the most significant future uses of Big Data in the judicial system. Information produced by the Data Analytics group in JUS indicated that the majority of cases processed by JUS are relatively small; in fact large cases are the outliers in statistical terms. While some cases processed by JUS must be taken to trial, many small cases may go to trial where the outcome can be predicted in advance. Significant savings in settling those cases without having to go to trial might result.

While JUS might be obliged to take a case to trial on principle, regardless of the possible outcome, predictive analytics may offer the opportunity to avoid trial in many situations.

---

[2] Structured data is organized in a highly mechanized and manageable fashion which can be easily processed by a computer, such as stored in Excel spreadsheets; by comparison, unstructured data, such as text found in e-mails and text reports is raw and unorganized. Searching through unstructured data can be expensive and difficult.

V-3: Current Limitations with Prediction Models

There are currently limitations to the predictive analytics approach to case outcome prediction and/or settlement. Predicting the outcome of new legal cases is still an imperfect science because of limitations of the current information is available for inclusion. e.g.:

- Cases may be settled without going to trial and aren't available for inclusion in the database, making the data incomplete;
- Courts may not have decided enough similar cases to permit the statistical prediction of case outcomes or feature weights that are need to resolve the problem of small or biased samples;
- Algorithms that rely solely on assigning quantitative feature weights can be problematic because they are not sensitive to the particular context of a problem;
- The statistical algorithms used in the prediction models require sufficiently large data sets and, the more difficult the task, the more cases are needed to achieve accuracy;
- Text cases need to be represented in an appropriate form to enable machine learning; this is currently a largely manual process

These difficulties are likely to be overcome with time and, given an appropriate database of cases, statistical or symbolic machine learning[3] techniques will be used effectively to determine general rules for classifying new cases and predicting their outcomes.

One major impediment to predictive analytics faced by JUS and the Canadian legal profession is the expense of building a complete and accurate Big Data store of cases and precedents. The information must also be kept current for new legal decisions and appeal results. It is unlikely that such a project could be funded in the near future without the backing of a consortium of law firms, or a third party organization such as LexisNexis, or CanLII, which is supported by the Federation of Law Societies of Canada. Either of these organizations might be willing to fund the Big Data initiatives described above as something that would benefit all Canadian counsel, possibly provided on a pay per use basis or available by subscription. However, a detailed cost-benefit analysis would need to be performed before embarking on such a large project.

As a comparison, new regulations governing the accounting profession in 1999 forced the large accounting firms in the US and internationally to commission the development of a database containing information of all public and private companies, for use in determining possible conflicts of interest impairing auditor independence. Collectively, the firms engaged Sentinel,

[3] Symbolic Machine Learning is another term used for predictive analytics or modeling where patterns of data are identified using human readable terms and symbols as opposed to numbers.

an organization supporting brokerage firms, to augment and modify their existing database of public and private organizations, and associated systems tools to accomplish this objective.

V-4: Data Management and Analytics

Controlling the information that is captured in large datasets can be problematic and subject to legal or ethical restrictions including:

- Documents used as evidence that contain personal information;
- Third party sources of information, such as articles or agreements that may be subject to copyright laws preventing open disclosure or dissemination;
- Confidentiality agreements where open disclosure could cause harm to a third party;
- Content compliance with government policies and practices.

Content management and curation of a JUS Big Data site will be an onerous task. Data will need to be kept current, as well as in compliance with laws and policies. Fortunately, software tools are being developed to assist with this process in the form of Data Management Solutions for Analytics (DMSAs). Gartner Group describes a DMSA as "a complete software system that supports and manages data in one or many disparate file management systems (most commonly a database or multiple databases) that can perform relational processing (even if the data is not stored in a relational structure) and support access and data availability from independent analytic tools and interfaces." (Gartner Inc., 2015) Organizations offering these tools include traditional IT firms, such as Teradata, Oracle, IBM, Microsoft, SAP and HP). However, new organizations, such as Cloudera, MapR, Actian and Pivotal are competing with the leaders.

JUS is already accumulating and beginning to use sources of internal data that would form part of its Big Data repositories. Toundjer Erman, indicated that his objective was the "integration of information from all JUS systems that generate Enterprise Resource Management information in order to get a holistic view of all JUS operations." In parallel, there is a need to consider what the new operational landscape should look like, and then to generate new ideas by "looking through different lenses" and gaining new insights. This would include taking into consideration what other governments and public sector organizations are doing to use Big Data and technology to improve legal service processes and efficiency.

Ultimately, existing JUS data analytics information could be combined with other data for use in predicting how the legal environment will change. For example, will new legislation trigger more litigation, and what resources will need to be recruited or developed in JUS over a period of 3-5 years to respond to those predicted needs. Predictive Data Analytics can contribute to this analysis, but will require redevelopment of the current data architecture in the administration and resource planning areas to be more process driven.

V-5: Policy Development

Big Data offers an opportunity to contribute to government policy debates. One method for measuring public opinion and acceptability of government policy is through Internet search activity. For example, Twitter contains a huge public archive of popular sentiment containing ideas, opinions and debates on government policy issues that can be incorporated into policy decision-making, both in terms of identifying the requirement for potential policy development, as well as the need for possible policy adjustment when required.

While government agencies can follow what people are saying on social media tools such as Twitter, complex decisions usually require input from a lot of different sources, which requires integrating complex systems and data for better decision-making. Software solutions and Internet "listening tools" are becoming available that can monitor and track multiple social media channels and analyze trends, including public sentiment. Specific key words and search terms can also identify relevant content for further analysis using data analytics tools and software.

Facebook and Yahoo are using Apache Hadoop for this purpose, while Amazon Web Services also has offerings in this area. Other tools such as "Social Harvest" and "Pentaho" can be used to extract data from Twitter, Facebook, and other social media platforms and log this information to a variety of data stores. Statistics Canada and many other government agencies possess a wide range of data concerning the behaviour of Canadians as a direct result of citizen interactions with government online services. However, a government department that uses social media to try to identify and better understand the needs of Canadians might also be accused of spying on its citizens in order to supress potential resistance.

The use of Big Data for policy development raises new moral and ethical issues for policy makers. Using predictive analytics and probability theory to predict what the general population might do in the future, as opposed to what they have done in the past could contribute to the policy debate. However, results based on findings from a relatively small group of people might still contain errors. A risk is that Big Data predictions about individuals might punish people for their propensities, not their actions, thus potentially denying basic human rights. Predictive analytics used by police in the US has led to a reduction in certain crimes, but resulted in the targeting certain socio-economic or cultural groups. (Joh, 2014)

Section VI: Other Government Big Data Sources and Uses

Big Data offers a wealth of opportunities for other government agencies. Table 2 (below) shows a few of the areas in which Big Data is being exploited by other governments around the world.

Table 2: International Governmental Uses of Big Data

	US	UK	Australia	New Zealand	Singapore	Israel	Other
Predictive Policing	✓	✓	✓	✓	✓	✓	Spain
Informed Sentencing	✓	✓		✓			
Bail/Parole	✓	✓	✓			✓	
Fraud Detection	✓	✓			✓	✓	Canada
Health Care	✓	✓			✓	✓	Taiwan
Education	✓	✓		✓		✓	Korea, Canada
Public Works	✓				✓		Ireland, Philippines
Transportation	✓		✓	✓	✓		Sweden, Ireland
Infrastructure				✓	✓		
Economic Policy			✓				Japan, Germany, Canada
Environment				✓			Netherlands, Canada
Public Relations					✓		Japan, Hong Kong, China
Information Sharing	✓	✓			✓		Spain, Ireland, Japan
Government Resource Allocation			✓	✓	✓		Philippines, Germany

The growing use of Big Data by governments around the world is a very broad topic area, and the research and development of a comprehensive list of applications would be a very large undertaking. Therefore the foregoing chart is meant as a representative sample only.

VI-1: Sentencing and Parole

Big Data can have a big impact on Correctional Services and on the Criminal Justice system in general by informing sentencing and parole decisions in a variety of ways.

Data-centered, evidence based strategies can be used to divert as many people as possible toward alternative programs, either within or outside of prisons, possibly reducing prison crowding and lowering the likelihood of re-offense. The Attorney General of the United States says "[d]ata can [...] help design paths for federal inmates to lower these risk assessments, and earn their way towards a reduced sentence, based on participation in programs that research shows can dramatically improve the odds of successful re-entry. Such evidence-based strategies show promise in allowing us to more effectively reduce recidivism" (Leopold, 2014). Similar risk assessments can be used to inform bail and parole decisions.

These types of strategies are being used effectively in a variety of jurisdictions:

The State of Florida and the province of Quebec both use statistical programs to profile juvenile offenders and assign them to risk-specific rehabilitation programs. These programs have shown significant success in reducing recidivism (Perry, McInnis, Price, Smith, & Hollywood, 2013);

The US states of Pennsylvania and Tennessee and the Australian state of New South Wales require statistical analysis to be used in all sentencing decisions;

The cities of Baltimore, Philadelphia and Washington, DC, all use algorithms to predict the likelihood of re-offence by parolees, and plan parolee supervision accordingly.

Big Data can also be used at a higher level to inform overall sentencing guidelines; the US Sentencing Commission is currently studying the use of data-driven analysis to issue general (not individual) policy recommendations. These could include changes in recommended sentence length where historical data shows current measures to be ineffective.

VI-2: Policing and Security

Law enforcement agencies have a history of using profiling and data mining to identify potential threats and predict criminal activity: Big Data offers a variety of tools to augment this capacity.

DEPLOYMENT

Predictive analytics are being used in over sixty major cities across the United States to help law enforcement agencies predict areas of probable criminal activity, and to assign patrols

accordingly. These programs take into account times and locations of previous crimes, incident records, weather patterns, and historical and sociological information to create maps of "hot spots". Cities using these maps have reported decreases of between 10% and 40% in criminal activity as a result. Los Angeles also tweets daily "hot spots" to citizens, to increase vigilance.

CRIME PREDICTION

Predictive analytics can also be applied more narrowly, to identify individuals at high risk of committing crimes. Chicago has a program in effect that uses a "heat list", created by a complex algorithm using data from a wide variety of sources. Officers or letters are sent to the homes of people on this list, to offer social services such as job training, or tailored warnings of increased penalties for certain crimes for people with particular prior convictions. The program has yielded positive results and is considered a success.

The U.S. Department of Homeland Security (DHS) and the Israel Security Agency (ISA) both have programs under development to detect terrorist attacks before they happen. DHS uses a Future Attribute Screening Technology to screen people for behavioural attributes associated with violent acts. Their Predictive Screening Project defines observable behaviours that precede a suicide bombing attack, and has shown promise in the testing phase. The ISA is investing in technology to convert unstructured data such as video and audio into a form that can be analyzed and used to produce real time alerts.

CRIME DETECTION

A third area of use for Big Data in policing is detecting crimes in near real time. This is being applied mainly to various forms of fraud, such as Medicare, securities, and bank fraud in the U.S. It is also being used in the UK to detect the misuse of prescriptions, and foreign bribery.

VI-3: Full Litigation Services

In 2013, Singapore launched a country-wide Integrated Electronic Litigation System for all litigation. iELS is accessible from anywhere through an internet browser, and has the following key functionalities (Braddell Brothers, 2015):

- Streamlining and re-engineering of high volume litigation processes;
- Information-based filing - Data capture (e.g. via XML and electronic forms) instead of only paper capture (e.g. document scanning), enabling the flexible re-employment of information as and when required;
- Active case management - Courts can pro-actively track and manage pending matters
- Litigation process management - Alerts and triggers designated to ensure that litigants do not miss critical deadlines;

- Electronic case file for lawyers - Lawyers have access to all relevant documents at any time and any place with an Internet connection, for the duration of each case;
- Integrated due diligence checks - Due diligence checks integrated with the electronic filing process, doing away with the need for subsequent back-room reconciliation;
- Court calendaring - Optimal assignation of court hearing days to be achieved with the syndication of date/scheduling information captured via information-based filing.

VI-4: Transportation

Intra and inter-city transportation systems (including both infrastructure and services) produce a vast amount of data from sources such as road sensors, bus GPSs, and ticketing systems that can be analyzed and used to increase the efficiency of services and allocate government resources. Some examples of foreign governments using these data to great advantage include:

- Swedish National Road Administration uses IBM systems to predict, control and optimize road traffic to improve air quality and reduce congestion. This resulted in peak-time road traffic congestion being dramatically reduced, air pollutants cut by up to 12 percent, and public transport usage increase significantly;
- The city of LA uses demand-responsive pricing for parking. Prices are based on data from parking sensors, surveys, weather forecasts, information about holidays, local business activities, etc.;
- The city of Dublin provides live road sensor and city bus GPS data to citizens, who can use it to plan their routes;
- Similarly, New Zealand uses predictive analytics to provide motorists with real-time information on traffic patterns via Variable Message Signs, in operation on highways across the country. These signs also display messages about accidents and road closures and conditions.

VI-5: Health Care

A large variety of health related data exists (patient records, genome information, successful/unsuccessful trials, hospital records etc.). By combining this data for analysis, variants of a disease can be identified, as well as subsets of patients who would benefit from different treatment plans. Following up with these groups could lead to better outcomes for the patients, and greatly advance the research, although this can be difficult if information is anonymized or de-identified. (President's Council of Advisors on Science and Technology, 2014)

Some examples of Big Data currently being used in the health care field include:

- New Jersey uses medical billing data to map out hot spots where there are the most complex and costly healthcare cases, as part of a program to lower healthcare costs

- The UK Food Standards Agency uses Twitter data to predict outbreaks in real time (often weeks before other methods);
- In Singapore, hospitals are using predictive analytics to predict relapses;
- Taipei Medical University analyzes and monitors performance across all hospitals.

VI-6: Economics

Reliable information about the current state of the economy is extremely important in making monetary policy decisions. Big Data can provide this information by predicting a wide variety of econometrics. For example, there are a variety of leading and lagging indicators of overall unemployment in a jurisdiction, such as automobile downgrades and decreased grocery spending (leading), and increased foreclosures and vacation cancellations (lagging). This sort of analysis can be used for early warning, real time awareness, and real time feedback for public policies and programs. (Letouze, 2012)

The Bank of Canada has suggested using existing monthly indicators in combination with big data to predict GDP growth before official quarterly National Accounts data are released providing more timely and accurate metrics to inform monetary policy decisions. (Armah, 2013)

VI-7: Education

With the advent and increasing popularity of online learning, there are new sets of data available about how and what students learn, including responses to various new techniques and modes of delivery. Research into these data could yield great benefits to the field of education, including identifying what skills taught at which points in childhood, leading to better adult performance in certain tasks. Learning management systems (for use in actual classrooms) are also becoming more popular, and are adding to the available data. (President's Council of Advisors on Science and Technology, 2014)

Student data can also be used to identify and respond to student having educational difficulty. In 2012, Ontario's Ministry of Education identified 14,000 students across the province who had left high school with three or less credits needed to graduate. One year later, after a campaign to get them to go to summer school or take extra credit courses, 8000 of them had graduated. (Solomon, 2013)

Section VII: Big Data and Privacy in Government

VII-1: Privacy Laws

There is a complex matrix of laws, regulations and practices that arise from the possible use of Big Data in Government, and might affect its usage. The major international, national, and provincial laws are summarized in Appendix XI-3 However, many other sector specific privacy laws and considerations exist that may also come into play, depending on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the type of consent obtained from the data subject, etc. The following observations can be made about the legislation:

- There is no single law or practice governing data privacy; legislation exists at all levels of government creating a complex matrix of international, national and provincial laws that govern the use of personal information in Canada and abroad.
- Although similar in concept, privacy laws are not always aligned; some laws also have cross-border and extraterritorial reach.
- Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA.
- Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may be in conflict with the Privacy laws.
- Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes, and may therefore be difficult to apply.
- There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.

The primary legislation that impacts the use of personal data by the public sector and therefore the Canadian Government is the Canadian federal Privacy Act, although there are situations in which private sector and other legislation (e.g. PIPEDA) will apply to the uses of personal information by JUS and other Government departments.

VII-2: Recent and Pending Changes to Privacy Legislation

All governments are struggling with ways to keep their data privacy legislation current, relevant, and usable in light of the rapid technological developments. Of particular concern are

the new analytical tools that have the ability to mine data and analyze the ever-increasing data sources, and especially those that target personal information. Perhaps of even greater concern is the trend toward consolidation of existing databases into Big Data sources. The concentration of personal information from various sources adds complexity and risk. Privacy laws in the international community are far from static, and changes are likely to have an impact on Canadian laws and practices as these occur.

"As business systems and processes become increasingly complex and sophisticated, organizations are collecting growing amounts of personal information. As a result, personal information is vulnerable to a variety of risks, including loss, misuse, unauthorized access and unauthorized disclosure. Those vulnerabilities raise concerns for organizations, governments and the public in general." (AICPA/CICA).

Recent changes made to legislation could have significant implications for personal data privacy and the rights of Canadians. The specific aspects of these laws are presented in Appendix XI-4. The laws include:

- Bill S-4 The digital Privacy Act
- Bill C-13 Protecting Canadians from Online Crime Act
- Bill C-51 Investigative Powers for the 21st Century Act (aka the "Anti-Terrorism Act")

VII-3: Legislative Restrictions, Guidelines and Safeguards Regarding Government Use of Personal Information

An increasing amount of information is available from the Canadian Government through its "Open Government" and its other initiatives; this trend is likely to continue. At the same time, controls have been established to try to ensure that personal information is only made available to those who are authorized to access it.

ACCESS TO INFORMATION AND PRIVACY PROGRAM (ATIP)

Systems are controlled and information is subject to review under the requirements of the Access to Information Act and the Privacy Act before being released. The ATIP program also permits citizens to determine what information government holds about them, and provides them with the ability to correct the information if it is inaccurate.

Government procedures also exist surrounding the handling of personal information by its departments and agencies. Guidelines issued by the Treasury Board include a Directive requiring the performance of an extensive Privacy Impact Assessment (PIA) before implementing or changing government systems, or altering the manner in which they process information. The PIA includes guidelines for the assessment of privacy implications before entering into contracts or making outsourcing decisions.

THE STATISTICS ACT

The Statistics Act permits StatsCan to enter into contractual agreements to share confidential information with other government departments under specific conditions:

1) Information can be shared with the statistical agencies of provinces and territories for statistical purposes if:
    a. The data subjects were notified at the time of data collection;
    b. The provincial agency has the authority to collect the information on its own; and
    c. The agency's confidentiality protection requirements are substantially the same as those of Statistics Canada.
2) Where information is collected jointly by Statistics Canada and any federal and provincial government department, municipal government or other incorporated body such as an association or university, and where data subjects are notified in advance of intention to share the data, and are given the opportunity at the time of data collection to refuse to allow their information to be shared.

The OPC has also highlighted the existence of other laws that supplement, but do not necessarily supersede, the Privacy Act and PIPEDA and which provide Canadians with additional protections for their personal information:

"Several federal and provincial sector-specific laws include provisions dealing with the protection of personal information. The federal *Bank Act*, for example, contains provisions regulating the use and disclosure of personal financial information by federally regulated financial institutions.

Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals." Source (Office of the Privacy Commissioner of Canada)

Therefore, many substantial controls do exist over the internal use of personal information by government departments and agencies.

VII-4: Implications for The Department of Justice

While privacy concerns are already one of the considerations that counsel are responsible for taking into account when using or posting any kind of legal reference document to Justipedia or

other data repositories, determining which jurisdiction governs personal information is becoming much more complicated as information is gathered and/or transferred across legal jurisdictions and co-mingled in Big Data stores or linked with other information sources. It is also easy to lose track of the origin of the data over time, and especially if the organization operates across Canada or captures information on the Internet. Maintaining data accuracy and responding to citizen's information requests becomes problematic. Courts around the world are struggling with data ownership and the determination of which laws will apply.

Most of the Big Data that is to be used by JUS is likely to consist of legal precedents and opinions, or possibly large quantities of evidence submitted in a court case to be analyzed using eDiscovery tools. Therefore, although there may be a few exceptions, (e.g. criminal records), JUS appears unlikely to capture and use a significant amount of personal Big Data, other than where it uses personal information contained in other government databases (e.g. StatsCan) for analytical purposes, and usually in aggregated form. However, the department may use personal information of its own staff members to assess efficiency and productivity of the various department functions. Also, while Government departments are primarily concerned with compliance the Privacy Act, PIPEDA may also apply to aspects of litigation proceedings, depending on the context, when personal information captured in connection with litigation involves commercial organizations or is carried out in the course of commercial activities.

Regardless, JUS is likely to be involved in legal actions or discussions surrounding the use of personal data by other Government departments, and some of the evidence that it collects which includes sensitive or other personal information must be kept private. In these cases, JUS lawyers will need to be respect their obligations under PIPEDA by ensuring that any personal information collected, used or disclosed in connection with any anticipated or actual litigation (or any other use) needs to be done either with the consent of the individuals, or must otherwise meets one of the applicable exceptions to the knowledge and consent principles of PIPEDA or the Privacy Act.

Section VIII: Privacy Concerns - Big Data and Government

The issues raised by the establishment of Big Data sources are not necessarily new, but relate to the difficulty of managing and protecting such large banks of information. Also, the sheer volume of the data held by government – both collectively and about each individual – creates the concern that profiles of individual characteristics and behaviour can be established that are quite complete and accurate. The application of predictive analytics to that information could permit the prediction of future trends and behaviours of both societies and individuals. While this might have benefits, there is a darker side to the existence of mass stores of personal data if the capability was misused.

The main concerns, discussed further below, are likely to be in four broad areas:

- Big Data Management and Security;
- Individual Consent regarding permitted uses of the personal information;
- Transparency and government disclosure of how data is collected, stored and used; and
- Lack of trust in government.

VIII-1: Big Data Management and Security

Large electronic sources of personal information can have significant value to those with less honourable intents. Once accessed, huge amounts of information can be rapidly transferred and stored inexpensively and with relative ease, attracting theft for monetary gain or extortion where personal exposure might have adverse impacts for both individuals and governments. The more attractive the information, the greater the difficulty to protect against data breaches by sophisticated hacking communities or tools – both in state-sponsored or private hands.

The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. This could involve greater risk of misuse in the event of a data breach, and eventual misuse for identity theft or fraud. The risk to government and individuals must be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

The demonstrated ability of hackers to overcome the security of government websites (e.g. recent attacks by Anonymous on Canadian Government web sites) and the perception that personal information is at risk of being disclosed or used fraudulently undermines public confidence in the safety of having their personal information in government data repositories.

"Each of the Canadian Privacy Statutes contains safeguarding provisions designed to protect personal information. In essence, these provisions require organizations to take reasonable technical, physical and administrative measures to protect personal information against loss or

theft, unauthorized access, disclosure, copying, use, modification or destruction. These laws do not generally mandate specific technical requirements for the safeguarding of personal information." (Piper, 2015)

Somewhat surprisingly, there are no prescribed standards for implementing security controls to protect personal information; rather it is left up to organizations to use their own judgement to determine what is appropriate. PIPEDA and the B.C. and Alberta privacy acts only "require organizations to take reasonable steps to safeguard the personal information in their custody or control from such risks as unauthorized access, collection, use, disclosure, copying, modification, disposal or destruction." (Office of the Privacy Commissioner of Canada, n.d.)

Reasonable safeguards include several layers of security, including, but not limited to risk management; security policies; human resources, physical and technical security; and business continuity management. The reasonableness of security arrangements adopted by an organization must be evaluated in light of a risk assessment including a number of factors, such as the sensitivity of the personal information; the foreseeable risks; the likelihood of damage occurring and the resulting harm caused; the medium and format of the storage method, and the cost of putting preventative measures in place.

VIII-2: Consent to Use Personal Information

The so-called "secondary use" of personal data - i.e. the use of data that has been provided for one purpose for other purposes - is a growing problem in the digital world, and in the Big Data world in particular. There is also a grey area between what information might require explicit or implicit consent for its use. The rules surrounding the requirement for consent and the use of personal information is clearly laid out for the private sector in PIPEDA, but Big Data will create broader issues for the public sector as well.

In the past, some of this data was considered to have been provided with the individual's implicit consent that it would be used in accordance with disclosures made by organizations. However, legislation covering the collection of most personal data collected by private organizations in Canada now requires explicit consent for use in accordance with specific terms. Any proposed secondary use for other purposes isn't generally permitted unless the use is disclosed at the time of collections. This is especially true in situations regarding the use of one of the sensitive categories of information (see Appendix XI-6).

Subject to legal interpretation, The Privacy Act might provide the government with more flexibility in its use of information provided to its various departments in the normal course of business, including the sharing and exchange of this information between government departments in the form of a Big Data repository, so long as the information is adequately protected from improper access or uses. Such use is already being made for research purposes

(e.g. by StatsCan). Sections 7 & 8 of the Privacy Act appear to cover this use. However, in the future expansion of Open Data and Big Data, where information is spreading out in many directions, it might be more difficult to determine whether information is being used in ways that don't require some form of additional consent or opt-out capability, and there may be unintended consequences. The standard form of consent or notification provided by the government will probably have to be worded very carefully at the front end of the process, and the back end of the process will require some form of careful review to ensure that the information is not being used outside of legal boundaries.

VIII-3: Transparency and Government Disclosure

The 2014 OPC survey reported, "The vast majority (89%) of those who had heard something about government surveillance activities agreed that surveillance or intelligence gathering agencies should have to explain their activities to Canadians." (Phoenix Strategic Projections Inc, 2014)

In 2000, the Canadian Government began to create its first Big Data repository, which became known as "Big Brother". The database included information on the addresses, education, marital status and ethnic origin of Canadians. It also tracked a person's employment and social assistance history, and their income tax records. Plans to implement the database were shelved at the time due to concerns expressed by the OPC and in Parliament, and also because of the volume of public requests to see their personal information contained in the database. (CBC News, 2000)

The concerns of the Canadian public in this area remain today. A conclusion of the 2014 OPC survey was that "The majority of Canadians are not confidant that they have enough information to know how new technologies might affect their personal privacy." This would likely extend to the enhanced use of Big Data by government. "Canadians expressed varying levels of comfort with different ways in which government departments and agencies, including intelligence gathering organizations, could collect or share their personal information." (Phoenix Strategic Projections Inc, 2014)

Only about half of the OPC survey respondents felt that:

- They had a good understanding of what the Government did with personal information that it collects;
- They were confident that the government would take their concerns about handling of their data seriously;
- They were confident that personal information shared with government would not be misused, lost or stolen.

VIII-4: Data Breaches and Trust in Government

While there are no statistics regarding trust in the Canadian government to protect personal information, numerous highly publicized data breaches have occurred in Canada over the past few years, and the numbers have grown substantially:

"The federal government reported breaching the privacy of individuals more than 5,000 times last year — an all-time high, according to new figures. The data are only for six departments, so the 5,237 privacy breaches they reported in 2014 are likely just a glimpse at what happened across government. Even so, the figure is almost as many as had been reported in the previous 11-year period, including instances where a taxpayer's or organization's information was incorrectly released, lost or compromised." (Press, 2015). Public awareness of attacks on government has increased too with the recent highly publicized attacks on Government web sites by "hacktivist" groups, such as Anonymous.

Canadians are waking up to the possible uses of their personal information by government agencies. The December 2014 OPC survey found that "56% of Canadians have some awareness of surveillance and intelligence gathering activities." "Roughly half (49%) of Canadians have seen, read, or heard something about surveillance or intelligence gathering activities for the purposes of national security in the past year or so." (Phoenix Strategic Projections Inc, 2014)

The heightened awareness of Canadians is likely a result of the recent publicity of government surveillance and information sharing programs through public revelations by Edward Snowden and the debate surrounding Bill C-51 (now the Anti-Terrorism Act) and its potential implications for the privacy of personal information. 78% of those polled in the OPC survey said they were either very (44%) or somewhat (34%) concerned about law enforcement and security agencies collecting their personal information for government surveillance purposes.

VIII-5: Big Data Privacy Controls

While the use of Big Data and related technologies can create significant privacy concerns as highlighted above, some of the technologies available now also permit the implementation of sophisticated controls to protect individual rights of citizens by regulating how Big Data technologies are used. Examples of these controls include:

- Use of methods for the "tagging" of data to ensure use is restricted to the purposes for which it was collected or generated;
- Implementing purpose-based or user-based controls according to the permissions and restrictions established for this data, including access controls;
- Tracking user access to data and the purposes for which it is used;

- Implementing algorithms that provide alerts regarding inappropriate access and possible uses.

While the use of specific information by JUS may not raise broad privacy concerns, other information available to government agencies may cause issues when data is aggregated or concentrated in electronic form, and especially when data is merged from multiple agencies. Regardless, there can be great benefits to merging and analyzing this information, such as:

- For research and public policy development regarding health, social, economic, national statistical trends;
- To demonstrate transparency and accountability of government; and
- To achieve public participation through engagement.

There is tremendous value in having broader access to this information for research, analysis, and policy development. Big Data is being used by the US Department of Justice to analyze medical billing records to detect Medicare fraud, and they are looking at similar Big Data sources for the detection of other frauds. (Scannell, 2015). The increased sharing of information across government departments also creates complex relationships and can result in difficulties surrounding disclosure and transparency about the use of the information. One such example is the Canadian Open Government Portal that is intended to provide "greater transparency and accountability, increase citizen engagement, and drive innovation and economic opportunities through Open Data, Open Information, and Open Dialogue" (Government of Canada, n.d.).

Achieving full openness while maintaining appropriate controls over data privacy may be mutually exclusive objectives requiring some compromises. Legal privacy objectives can often be achieved through "de-identification" or "anonymization" of data, but the more heavily data is neutralized in this manner, the less useful it can become. In addition to legal requirements for compliance, there are also ethical considerations and, while privacy and confidentiality are somewhat different concepts, contractual and other agreements regarding the possible use of information (e.g. copyright) may need to be considered.

Focus groups during the 3rd International Open Data Conference held recently in Ottawa identified several privacy concerns and issues around open data:

- The public sector collects a great deal of sensitive personal information. While individual sources of anonymized or de-identified information might not reveal the identity of a person, the use of multiple data points that link or connect to others may make it possible to connect or triangulate between unrelated data points, making it possible to identify individuals.
- The use of Census and national statistical information can be problematic, even if data is aggregated, since individuals can often be identified within small groups or communities. Locational data can sometimes involve the same risk as a personal identifier "key", such as a name or social insurance number.

- The potential to profile, target or discriminate against vulnerable people or groups might be possible through matching of open data sources with information gained from other private sources.

The concern exists that while government surveillance will be made easier for protection against terrorism and illegal acts. (Open Data Ottawa Privacy Conference Notes)

VIII-6: Forecasting Canadian Public Opinion on Privacy and Big Data in Government

According to a study conducted by the Office of the Privacy Commissioner (OPC) in December 2014, "Nine in ten Canadians expressed some level of concern about the protection of their privacy, with 34% saying they are extremely concerned (up from 25% in 2012)." Further, "Canadians increasingly feel that their ability to protect their personal information is diminishing. Seventy-three percent, the greatest proportion since tracking began, think they have less protection of their personal information in their daily lives than they did ten years ago." (Phoenix Strategic Projections Inc, 2014)

Canadians are therefore aware and concerned about the privacy of their personal information, and increasingly so. The primary focus of Canada's privacy programs has arguably been on the use of personal information in the private commercial sector, and the protection of this data through PIPEDA and its enforcement by the OPC. The same degree of knowledge or awareness of the Privacy Act and the permissions afforded by it to government doesn't seem to exist.

At the same time, recent legislative changes (see Appendix XI-4) and public revelations concerning clandestine government surveillance programs by Western governments, including Canada, have not likely helped to ease public concern. Some vocal members of the Canadian public, in particular, are questioning whether the extent to which the legislation is being implemented is commensurate with the need.

Anti-Terrorism Bill C-51, in particular, appears to be the subject of much concern. Daniel Therrien, the Privacy Commissioner of Canada, is responsible for the independent oversight of Canada's privacy laws and compliance. He recently submitted an article published in the Globe and Mail in which he said:

"In my view, Bill C-51, in its current form, would fail to provide Canadians with what they want and expect: legislation that protects both their safety and their privacy. As proposed, it does not strike the right balance.

The scale of information-sharing between government departments and agencies proposed in this bill is unprecedented. The new powers that would be created are excessive and the privacy safeguards proposed are seriously deficient." (Therrien, 2015)

The focus on the use of Big Data to track people and groups casts a negative image. The Commissioner's comments and position on information sharing are likely to create further debate and shape public opinion regarding government Big Data and data sharing between departments and with other governments. As sharing of Big Data information by government agencies becomes more commonplace, Canadians may become increasingly concerned about the possible uses, and react negatively.

Addressing the lack of awareness by Canadians of the way in which their personal information is being used may require greater emphasis on public disclosure to help reduce concerns. One recommendation made by the recent report to US President Obama suggests the implementation of a Consumer Privacy Bill of Rights based on the Fair Information Practice Principles. While this approach might help confidence in the private sector, a broader "Citizen's Bill of Rights" might be more appropriate to help renew the trust in government to protect personal information in the face of the expanded use of Big Data. One of the Key Informants, Howard Deane, from the Consumers Council of Canada, expressed the view the level of trust might be elevated if the government was more transparent regarding how personal information that makes its way into their hands will be used (i.e. limits on use), and what protections will be put into place around Big Data to avoid its misuse.

There are also ethical and moral questions about how Big Data might be used by government, or disclosed to others for possible misuse. There is a difference between government predicting and disclosing broad statistics about crime and cancer rates on a macro scale and using the data to focus in on individuals. The more granular the information becomes, the more organizations might be tempted to use the information in negative ways.

In his Globe and Mail article, the Privacy Commissioner indicated that the new legislation would "provide 17 federal government agencies with almost limitless powers to monitor and profile ordinary Canadians, with a view to identifying security threats among them. The end result is that national security agencies would potentially be aware of all interactions all Canadians have with their government. That would include, for example, a person's tax information and details about a person's business and vacation travel." (Therrien, 2015)

Public opinion is often difficult to predict because it often varies by culture, and is subject to "trigger" events that cause rapid shifts - e.g. The Edward Snowden disclosures surrounding government surveillance involved such a shift. Other than the OPC survey conducted by Phoenix Strategic Projections Inc., there appear to be few detailed Canadian surveys and public opinion polls that specifically address this topic in detail, but recent studies done on public

perceptions and opinions in the US and EU confirm that people believe that the privacy and security of their personal information is at risk, as is their ability to keep their information confidential in such an open world. However, there are a number of recent international studies that support this view.[4]

A Welcome Trust study in the UK found that focus group participants distinguished between acceptable types of government uses of personal data according to the following factors:

- The Government identifying needs, planning resources and services, and allocating funds;
- Prevention and detection of crime and, including terrorism;
- Identifying social/population trends and statistics;
- Unearthing dishonesty (e.g. fraudulent benefit claimants and tradesmen)

While there was a general awareness of data collection by both government agencies and companies generally, the Welcome study found that the public views of the collection and use of personal data could be summarized as follows:

- The public consider the collection and use of personal data to be a big issue;
- When asked, the public are ostensibly opposed to any form of data use and collection by government and companies;
- In practice, the public consider there to be no alternative to sharing personal information with government and companies in the modern world and expect this to increase in future;

A significant proportion of the public expected to feel less comfortable about sharing personal data in future.

VIII-7: Summary

The 2014 study commissioned by the President of the United States regarding Big Data and Privacy included the conclusion that:

"Although the use of Big Data technologies by the government raises profound issues of how government power should be regulated, Big Data technologies also hold within them solutions that can enhance accountability, privacy, and the rights of citizens." "Responsibly employed, Big Data could lead to an aggregate increase in actual protections for the civil liberties and civil

[4] - PEW Research Study - Public Perceptions of Privacy and Security in the Post-Snowden Era - November 2014
- White House Study - Big Data and Privacy Review - May 2014
- EU Byte Study - Report on public perceptions and social impacts relevant to Big Data - March 2014
- Eurobarometer Report - Attitudes on Data Protection and Electronic Identity in the EU - June 2011

rights afforded of citizens, as well as drive transformation improvements in the provision of public services. " (Report to the Executive Office of the President)

It remains to be seen how the use of Big Data will translate into privacy concerns and the reaction by Canadians to the use of their personal information in Big Data repositories going forward. The level of trust in government, along with knowledge of why data is being collected and how it will be used also appear to be significant Issues, judging from recent public reaction to Bill C-51. There will likely be a need for programs to educate the public about these uses, and to promote the benefits, in order to establish a level of confidence and trust in the process, and to prevent a negative backlash such as occurred with the "Big Brother" database proposal in 2000.

Section IX: Major Findings and Conclusions

While Big Data is reforming many aspects of the world in which we live, the earliest successful models have been built on large databases of structured quantitative data, because this type of information is more easily and readily interpreted by binary computer logic. Although there are some exceptions, much of the information generated by the legal community or used in trials is unstructured data – e.g. reports, e-mails, and legal precedent cases. The technology-enabled tools required to analyze these files are complex and will take time to develop and refine.

Despite this, considerable progress has already been made in the development of IT tools and infrastructures to support innovative uses of Big Data elsewhere in the legal profession. The marketplace has proved to be very lucrative, and many new players have entered the field with significant financial backing and resources. New innovative solutions are emerging that offer the promise of both competitive advantages and cost efficiencies to those who adopt them.

JUS Query - Government departments and agencies continue to accumulate a wealth of data. At a time when governments are being asked to do more with less while providing new services to citizens, what might a "Big Data Strategy" for the Government of Canada look like?

IX-1: Possible Big Data Strategy

JUS will find itself increasingly to be in competition with other organizations in the legal community as they make investments in these new technologies to increase their efficiency and effectiveness in the courtroom, and the Department will find it necessary to make similar investments, since all parties will need to move forward to remain competitive and cost-effective, and to avoid being placed at a disadvantage. The strategic decision to be made by JUS is whether it should position itself as an early adopter of the technology, or be satisfied to be a "fast follower". The other decision will be how it should invest its limited resources to achieve its strategic objectives – i.e. what should the priorities be?

The one overarching conclusion that can be derived from the study is that large legal organizations that fail to plan for the implementation of these new technologies are likely to find themselves at a significant disadvantage from a competitive and cost-effectiveness standpoint. Donald Wochna, chief legal officer of Vestige Digital Investigations, was quoted in *Law Technology News* as saying: "Big Data in general, and predictive data analytics in particular, are the potential holy grail in the practice of law."

JUS Query - How can the Department of Justice adopt Big; Data for its own needs?

IX-2: Possible Uses of Big Data and Predictive Analytics in JUS

The possible uses of Big Data by JUS, and the implications thereof, include:

POSSIBLE APPLICATIONS OF BIG DATA BY JUS

The primary applications of Big Data analysis in the Department of Justice are expected to be the use of:

1) eDiscovery software tools to analyze and refine information in large databases of relevant documents for the production of evidence to be used in trials.
2) Predictive analytics and artificial intelligence to predict the outcome of cases based on a Big Data repositories of precedents and legal opinions, such as Justipedia, iCase or SAS, which might ultimately be used to reduce time spent and effort devoted to settling cases or taking them through the trial process.
3) Data analytics techniques to analyze large databases of JUS operational statistics, with a view to improving individual performance and the overall productivity and cost-efficiency of the Department and, in future, to proactively position department resources to address emerging trends.
4) Data analytics to predict environmental trends, based on both internal (e.g. StatsCan) and external information (e.g. social media) that might be used to respond to the need for changes in government policies.
5) Automated tools for early identification of legal risk associated with individual legal cases, and to manage risk throughout the trial process.
6) Automated tools to measure both individual performance and compliance with department and professional policies, procedures and standards and, in summary form, for management reporting of department performance and legal risk management.

SOME CONSIDERATIONS SURROUNDING BIG DATA IMPLEMENTATION

Lawyers who have already been exposed to the use of eDiscovery, predictive analytics and other advanced technology tools are recognizing some of the implications as well as the potential opportunities of working with advanced technologies and applying these tools to large repositories of relevant data. However, this is still a relatively new concept for many in the legal profession, and so it is difficult for them to know where to begin with plans for implementation. The following issues will need to be considered:

7) The legal world is definitely headed down a path where sophisticated technologies (e.g. Big Data and Predictive Analytics) will play an increasing role. Legal organizations that

fail to keep up will eventually find themselves to be at a competitive disadvantage in terms of managing litigation cost and achieving success in the trial process.

The implementation and adoption of complex new technologies can be a significant undertaking in large organizations such as JUS, and therefore takes considerable time (i.e. years) to accomplish. Advance planning is therefore critical to ensure that resources are available, and the implementation is a success.

8) Implementation of the new technology tools and processes will require a strong change management program, based on the inherent resistance of people to significant change. The input we received, both in JUS and externally, is that such a program is likely to be required in order to achieve widespread adoption of new technologies, and also systems that attempt to measure individual performance more closely.
9) Significant investment will be required over many years, in money and human resources to remain current with the external legal marketplace to avoid falling behind. This is especially true with respect to the use of Big Data and predictive analytics technology where there have been, and will be, significant developments in the legal community
10) JUS may not have access to the financial, human, and other resources required to move down all the emerging technology paths at once. The various options will need to be prioritized based on the projected cost/benefit before proceeding with any plans to implement Big Data, and considered as part of an overall departmental strategy.
11) Successful implementation is likely to depend on the ongoing commitment of JUS management to invest in the change, and to implement the tools required over a protracted period of time.

POSSIBLE COST EFFICIENCIES TO BE DERIVED FROM BIG DATA AND ADVANCED TECHNOLOGIES

The implementation of advanced technologies can be very expensive and disruptive to JUS, but the organization is likely to achieve both quality and cost-effectiveness improvements as a result. The following possible benefits were highlighted during our research:

12) Productivity and quality improvements would result advanced expert search technology and litigation support tools to better research information and relevant evidence;
13) Possible process and efficiency improvements could be achieved in JUS administration and operations;
14) Productivity could be improved through the use of advanced analytics to allocate litigation resources by predicting forward demand and adjusting supply of legal resources accordingly.
15) Costs might be reduced through the ability to use Big Data and predictive analytics to predict case outcomes and resolve cases without going to trial, possibly through the use of a dispute resolution process.

16) Access to internal and external Big Data sources would permit better policy decisions.

IS JUS POSITIONED TO TAKE ADVANTAGE OF NEW TECHNOLOGIES?

While the main purpose of this study was to look forward at possible uses of Big Data in JUS, the current uses of Information Technology in the Department was also reviewed. This is important as a starting point because the transition to the use of Big Data and predictive analytics in most large organizations relies on having a relatively strong base of technology on which to build. However, some technical and organizational restructuring may be required in order to move forward with more sophisticated technology programs.

JUS appears to have a variety of available technology tools, but there is some question as to how extensively these tools have been accepted and are being used by Department staff. In addition some of the key systems (e.g. iCase) are aging and in the process of being replaced with Government standard tools, although implementation is just beginning.

Regardless, the conclusion is that:

17) No serious technology impediments were identified that would prevent JUS from moving forward with Big Data projects.

JUS Query - Are there promising practices in other countries and departments worth emulating? Where and what are they?

PRACTICES IN OTHER COUNTRIES AND DEPARTMENTS

Sections D, E, and F of the report go into considerable detail about findings in this regard. The findings were mixed. Although there are some promising developments in other countries or departments that could be followed up, or developments to be followed, there don't seem to be any "magic bullets" at this time. However, there appears to be steady progress, and suppliers of technology in this area are making considerable investments.

18) Carrying on a "watching brief" of activities in other countries, while preparing to move forward as a clearer path emerges, might be an appropriate strategy for JUS.

JUS Query - What potential regulatory mechanisms, other than the traditional Organization for Economic Cooperation and Development (OECD) data protection principles, exist that could protect privacy as Big Data analytics become more widely used in the public and private sectors? Please do not limit the options of regulatory mechanisms to traditional modes of government regulation, but include any market mechanisms, technological mechanisms, incentives, social innovation, professional regulatory mechanisms, and private initiatives that could operate in this regulatory space. Please provide specific examples of these mechanisms.

IX-3: Government Big Data, Data Privacy and Public Opinion

PRIVACY LAWS

A complex matrix of laws, regulations and practices impact the possible use of Big Data in Government:

19) Canada is one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.
20) Many laws that were established before the proliferation of information technology and the age of Big Data did not anticipate the possible aggregation and uses of personal information, both for positive and potentially negative purposes.
21) There is no single law or practice governing data privacy; legislation that exists at all levels of government – a complex matrix of international, national and provincial laws exist that govern the use of personal information in the private and public sectors in Canada and abroad.
22) Although national laws are similar in concept, privacy laws are not always aligned; some also have cross-border and extraterritorial reach. Different laws govern the privacy of personal information in the Public and Private Sectors – e.g. The Privacy Act and PIPEDA There appears to be no reconciliation of the various laws governing privacy, so decisions regarding the application of the various laws are frequently resolved in the courts.
23) Other laws impact possible uses of personal data – e.g. The Canadian Charter of Rights and Freedoms and The Anti-Terrorism Act and must be considered and may also be in conflict with the Privacy laws. Other laws and agreements must also be considered – e.g. copyright laws and contract laws may govern the use and disclosure of personal and other data.
24) Jurisdiction of data privacy laws and the determination of which applies depends on many factors, such as the type of personal information, the location from which it was collected, and where it is processed and stored, the consent obtained from the data subject, etc.

JUS Query - What, if any, unique features or specific applications of Big Data analytics are likely to challenge Canadians' expectations of privacy in the short and medium term?

IX-4: Other Challenges to Privacy in a Big Data World

DATA SECURITY AND BREACHES

25. The greater the concentration of personal data in large or linked datasets, the greater the potential exposure if information is released. Government programs to expand access to data through the Internet also create additional points of potential entry for breaches to occur.

While data contained in Big Data repositories is unlikely to be released in volume, the ability to access a wide range of views of various information sources through available portals, and potentially to use sophisticated search capabilities to retrieve information can be causes for concern if the appropriate level of security and controls aren't in place.

26. The risk to government and individuals will need to be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions.

PUBLIC OPINION AND REACTION TO GOVERNMENT BIG DATA

One of the ultimate factors impacting Public Opinion and Reaction will likely be the level of trust in government. "Bad news" stories regarding events about government surveillance and data breaches can create an environment where citizens become very concerned about their information and negative public opinion goes "viral".

The laws surrounding the sharing of personal information, and the extent to which this can occur between government agencies and departments are unclear when it comes to sharing Big Data repositories and information. Public surveys in various countries have shown that the public are generally opposed to any form of data collection, use and sharing by government.

Government and organizations alike will need to deal with the issue of generally negative public reaction to the use of their private information. There are a number of factors, in particular, that might trigger a negative public reaction or, conversely, steps might be taken to mitigate a negative reaction from occurring.

27. The implementation of a Big Data repository by government is likely to require greater government transparency about the way in which government handles personal information in Canada, and a significant rethinking and restructuring of the ways in which personal information is protected in government hands.

JUS Query - What other options for moving forward would ensure adequate protection of Canadians from the negative implications of Big Data analytics?

28. There will likely be a need to re-examine and revise the various laws affecting personal data privacy in Canada, and especially as government and other Big Data projects are brought on stream. In this regard, any changes to the legislation need to be forward thinking regarding emerging technologies (e.g. the laws need "to go where the puck is going to be" with privacy legislation, and not where the puck has been) otherwise laws will become quickly out-dated.

Individuals with legitimate access rights (e.g. government employees) who are able to download information can also be a source of concern if that information is lost or compromised. There are controls that can be put in place to partially guard against these sorts of occurrence, but they are generally expensive and cumbersome to implement.

Section X: References

AICPA/CICA. (n.d.). *http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/PRIVACY/GENERALLYACCEPTEDPRIVACYPRINCIPLES/Pages/default.aspx.*

Armah, N. A. (2013). *Big Data Analysis: The Next Frontier.* Bank of Canada.

Braddell Brothers. (2015). *Singapore Litigation Procedure*. Retrieved from Braddell Brothers: http://braddellbrothers.com/litigation.html

Brown&Ehrenreich. (2015, July 13). Can Big Data and Privacy Coexist?

CBC News. (2000). *Ottawa breaks up 'Big Brother' database*.

Dwoskin, E. (2014, August 22). Can Big Data Improve Medical Diagnoses? *Wall Street Journal*.

EDRM. (2015, 1 1). *www.edrm.net.* Retrieved 6 9, 2015, from EDRM.net: www.edrm.net

Gartner Group. (2014). *Magic Quadrant for E-discover Software.* Gartner Group.

Gartner Inc. (2015, June 14). *IT Glossary*. Retrieved from Gartner Group: http://www.gartner.com/it-glossary/big-data

Google. (n.d.). *Google flu trends*. Retrieved from goog.org flu trends: http://www.google.org/flutrends/

Government of Canada. (n.d.). Retrieved from Canadian Open Government Portal.

IBM. (2015, June 16). *What is Big Data*. Retrieved from Big Data at the Speed of Business: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Joh, E. E. (2014, February). *Policing By Numbers: Big Data and the Fourth Amendment.* Retrieved from Washington Law Review: SSRN: http://ssrn.com/abstract=2403028

Krasnyansky, A. (2015, January 29). *Meet Ross, the IBM Watson-Powered Lawyer.* Retrieved from PFSK Labs: http://www.psfk.com/2015/01/ross-ibm-watson-powered-lawyer-legal-research.html

Leopold, G. (2014). AG Says Big Data Can Reform Sentencing Rules. *HPC Wire*.

Letouze, E. (2012). *Big Data for Development: Challenges and Opportunities*. New York: UN Global Pulse.

Library and Archives Canada. (n.d.). Legislative Restrictions: Records of the Government of Canada.

Library of Parliament Research Publications. (2014). *Lgislative Summary of Bill S-4*. (L. o. Parliament, Producer) Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?ls=s4&Parl=41&Ses=2&source=library_prb&Language=E#a1

Library of Parliament Research Publications. (2015). *Legislative Summary of Bill C-51: Investigative Powers for the 21st Century Act*. Retrieved from http://www.parl.gc.ca/About/Parliament/LegislativeSummaries/bills_ls.asp?Language=E&ls=c51&Parl=40&Ses=3&source=library_prb

Microsoft acquires Equivio. (2015, January 20). Retrieved from blogs.microsoft.com: http://blogs.microsoft.com/blog/2015/01/20/microsoft-acquires-equivio-provider-machine-learning-powered-compliance-solutions/

Office of the Privacy Commissioner of Canada. (n.d.). *A Privacy Handbook for Lawyers: PIPEDA and Your Practice.* Government of Canada, Office of the Privacy Commissioner.

Office of the Privacy Commissioner of Canada. (n.d.). *https://www.priv.gc.ca/resource/fs-fi/02_05_d_15_e.asp.*

Office of the Privacy Commissioner of Canada. (n.d.). *Securing Personal Information: A Self-Assessment Tool for Organizations*.

Open Data Ottawa Privacy Conference Notes. (n.d.).

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.* Rand Corporation.

Phoenix Strategic Projections Inc. (2014). *2014 Survey of Canadians on Privacy.* Canadian Federal Government, Office of the Privacy Commissioner.

Piper, D. (2015). *Data Protection Laws of the World.*

President's Council of Advisors on Science and Technology. (2014). *Report to the President: Big Data and Privacy: a Technological Perspective.* Washington, DC: Executive Office of the President.

Press, J. (2015, March 22). Federal government privacy breaches soar to record high. *Ottawa Citizen*. Ottawa, Ontario, Canada.

Report to the Executive Office of the President. (n.d.). *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.*

Scannell, K. (2015, January 12). DoJ uses big data to crack Medicare fraud schemes. *FT.COM.*

Shaw, J. (2014, April). Why "Big Data" Is a Big Deal. *Harvard Magazine*.

Solomon, H. (2013, June 27). *How Ontario faces big data privacy challenges*. Retrieved from IT World Canada: http://www.itworldcanada.com/article/how-ontario-faces-big-data-privacy-challenges/47722

Therrien, D. P. (2015, March 21). *Without big changes, Bill C-51 means big data.* Retrieved July 2015, from Globe and Mail: http://www.theglobeandmail.com/globe-debate/without-big-changes-bill-c-51-means-big-data/article23320329/

Transparency Market Research. (2014). *eDiscovery Market Global Industry Analysis, Trends and Forecast 2014 - 2020.* Transparency Market Research.

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions.* University of St Andrews, UK.

Section XI: Appendices

XI-1: Current Industry Leaders in eDiscovery (Gartner Group, 2014)

- kCura markets the Relativity platform that supports collection, legal hold, processing, review, analysis and production of evidence. Relativity is sold through a wide range of service providers and hosting partners, and through a growing direct sales channel.
- FTI Technology, a separate business unit of FTI Consulting, offers both eDiscovery software and services. Its main Ringtail platform performs functions from processing to evidence production. The Attenex product, also offered by FTI, provides a combination of machine learning and visual graphics for ease of document review.
- Recommind is known for its predictive coding technology, and supports all stages of the EDRM. Axcelerate eDiscovery can perform legal hold, collection, processing, review, analysis and production of documents, with Early Case Assessment and predictive coding capabilities.
- ZyLAB has an integrated solution supporting all stages of the EDRM. ZyLAB Intelligent Information Governance is used for file analysis and classification. Its eDiscovery technology architecture is horizontally scalable and can handle large datasets.
- HP's Autonomy eDiscovery tool supports the full process of EDRM. Their self-service eDiscovery OnDemand model is part of an ongoing product development initiative that addresses the market shift toward organizations that want to bring eDiscovery in-house. The product has a wide range of stakeholders ranging from IT users to in-house general counsel.
- Nuix's products include eDiscovery, Enterprise Collection Center, Web Review & Analytics, and Legal Hold. Its technology also extends to other related use cases, such as archive migrations, information governance and information security.
- Exterro provides products to support eDiscovery from identification through review. Its primary offering is the Exterro Fusion E-Discovery software suite, which is built on a single open platform. Exterro's Fusion Integration Hub allows integration of existing legal, eDiscovery and other information management systems.

XI-2: International Privacy Legislation

The original concept of data privacy was developed long before the explosion in the use of information technology could be envisioned. The impact of the new technologies used both in personal lives and in business is now apparent. The use of technology to access and manipulate personal data will continue, and is placing serious pressure on existing data privacy laws and practices around the world to keep up with the pace of change.

Political, geographical and cultural issues have made it difficult to adopt a single standard set of laws for data protection. Many different laws and regulations prescribe the privacy and treatment of personal information processed in Canada and in other legal jurisdictions. Most data privacy regimes include a range of seven to ten common principles. Those with a fewer number generally combine some of the principles, with a result that is largely the same.

The major forms of international legislation in place that prescribe the treatment of personal information from a data privacy standpoint are:

EU PRIVACY DIRECTIVE

One of the original, and arguably the strongest of the international privacy regimes, is the EU Directive (Directive 95/46/EC of the European Parliament and Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data) enacted by the European Parliament in October 1995. The EU Directive forms the basis for most national data privacy regimes in place around the world today. Only countries with privacy regimes in place that are deemed adequate by the EU are permitted to receive personal information from EU countries. The Canadian public sector Privacy Act and private sector "Personal Information Protection and Electronic Documents Act" (PIPEDA) and their application have made Canada one of the few countries accepted by the EU as being deemed adequate by the EU for such data transfers. This status speaks to the strength of Canada's privacy laws and needs to be preserved as it gives Canada an economic advantage over the many other trading nations who do not have the same status.

OECD GUIDELINES

The OECD Guidelines, issued in 1980 and revised in 2013, prescribe eight Basic Principles for National Application that align broadly with the EU Directive. The ten PIPEDA principles largely conform to the OECD standards and principles. The OECD principles follow:

1. Collection Limitation Principle

There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

2. Data Quality Principle

Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

3. Purpose Specification Principle

The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

4. Use Limitation Principle

Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except:

a) With the consent of the data subject; or

b) By the authority of law.

5. Security Safeguards Principle

Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.

6. Openness Principle

There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

7. Individual Participation Principle

An individual should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;

b) to have communicated to him, data relating to him:

i) Within a reasonable time;
ii) at a charge, if any, that is not excessive;
iii) in a reasonable manner; and
iv) in a form that is readily intelligible to him;

c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and

d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.

8. Accountability Principle

A data controller should be accountable for complying with measures which give effect to the principles stated above.

APEC PRIVACY FRAMEWORK

The Asia Pacific Economic Cooperation (APEC) Privacy Framework provides for a flexible approach to information Privacy protection across member economies to avoid the creation of unrealistic barriers to information flows. The framework contains nine principles that are similar to the EU Directive, OECD Principles and PIPEDA. Privacy enforcement relies on authorities from participating APEC economies, including the Office of the Privacy Commissioner in Canada.

XI-3: Canada's Public and Private Sector Privacy Legislation

Canadian privacy legislation is aimed separately at the private and public sectors, and there are important distinctions between the two: the private sector includes privately owned, non-government entities, while the public sector includes organizations that are owned and operated by the federal, provincial, and municipal governments. Some examples are:

- Educational institutions such as universities, colleges, technical institutes, school boards;
- Provincial and regional health care institutions, nursing home operators, hospital boards and subsidiary health corporations;
- Local governments, including municipalities, police services and libraries.

Perhaps the most important difference between PIPEDA and the Privacy Act is that the former provides certain guarantees regarding the collection and use of personal information collected by private sector organizations, setting the stage for possible legal remedies and actions in the event of improper use and/or disclosure, whereas the Privacy Act does not set the same limitations on use of the information, nor does it provide for specific actions or remedies by government in the case of misuse, disclosure or data breach.

PUBLIC SECTOR PRIVACY LAWS - THE FEDERAL GOVERNMENT PRIVACY ACT

The Federal Privacy Act, first enacted on July 1,1983, applies to all of the personal information that the federal government collects, uses and may disclose about individuals or federal employees, i.e. it sets out policy surrounding the Government's collection, use and disclosure of their personal information in the course of providing services (e.g., passports, pensions, taxes).

The Privacy Act also sets out how federally regulated public bodies can collect, use, and disclose personal information, as well as how individuals can ask to access and update their personal information Examples of federally regulated public bodies include the:

- Bank of Canada
- Canada Revenue Agency (CRA)
- Canadian Space Agency
- National Research Council Canada
- Statistics Canada
- Treasury Board of Canada

The Act also gives the federal government a wide range of powers surrounding their possible uses and disclosure of personal information, subject to certain controls. The ability to share personal information across government agencies appears to be facilitated by the provisions of the Privacy Act. For example, Section 8 of the Act provides for disclosure in accordance with legal agreements between federal government departments, provinces and territories, First

Nations councils, and foreign governments for the purpose of administering or enforcing laws. Recent legislation further enhances the capability of sharing personal information across government agencies. See Appendix XI-4)

The Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with the Privacy Act.

PRIVATE SECTOR PRIVACY LAWS - THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC ACT

Federal legislation governing personal data privacy in Canada is provided in the Personal Information Protection and Electronic Documents Act (PIPEDA), which establishes the manner in which private sector organizations can collect, use or disclose personal information while conducting commercial activities in Canada. It also applies to the personal information of the employees of federally regulated organizations, such as telecommunications companies, banks and airlines. However, PIPEDA does not apply to non-commercial organizations such as charities or not-for-profits or political parties and some non-commercial associations.

PIPEDA generally applies to:

- Private sector organizations carrying on business in Canada in the provinces or territories, when the personal information they collect, use or disclose crosses provincial or national borders (except for the handling of employee information).

- Federally-regulated organizations with commercial operations in Canada, such as airlines, banks, telephone or broadcasting companies, but including their handling of health information and employee information.

PIPEDA sets out the following ten principles, which are closely aligned with the EU Directive, OECD Principles, and the principles adopted by the CICA and AICPA as "Generally Accepted Privacy Principles" (GAPP) (Appendix XI-5). The following privacy concepts are covered in the PIPEDA principles:

1. Accountability;
2. Identifying purposes;
3. Consent;
4. Limiting collection;
5. Limiting use, disclosure and retention;
6. Accuracy;
7. Security safeguards;
8. Openness;
9. Individual access; and
10. Compliance.

As with the Privacy Act, the Office of the Privacy Commissioner of Canada is responsible for overseeing compliance with PIPEDA.

ROLE OF THE OFFICE OF THE PRIVACY COMMISSIONER OF CANADA

The Office of the Privacy Commissioner of Canada (OPC) advocates for the fundamental privacy rights of individuals through the establishment of an appropriate regulatory framework, and through the provision of independent oversight and monitoring of the application of PIPEDA to the private sector and the personal information handling practices of federal government departments and agencies to ensure compliance with the public sector Privacy Act.

OPC also acts as an ombudsman, working independently to:

- Advise individuals, government, businesses, and Parliament on emerging privacy issues;
- Investigate complaints and make recommendations based on findings; and
- Conduct audits under the two federal privacy laws; and
- Promote awareness and understanding of the protection of personal information.

PROVINCIAL LEGISLATION

Every province and territory has its own public sector legislation and these provincial acts also apply to provincial government agencies. Alberta, British Columbia and Québec have privacy legislation that is considered "substantially similar" to PIPEDA, so that the provincial act can be applied to private-sector businesses that collect, use and disclose personal information while doing business in those provinces. Ontario, New Brunswick, and Newfoundland and Labrador also have their own health care privacy legislation that supersedes PIPEDA in this area.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

Although provincial privacy laws are similar to federal laws, some important differences exist. For example, certain provincial privacy laws (e.g. Alberta) have special consent and transparency legislation that applies to organizations and/or their service providers who permit access to or disclose personal information to locations outside Canada. If this includes public sector bodies, it could create a conflict where information about an individual resident in a province might be shared with other governments, (e.g. in the case of suspected criminal, terrorist, or other activity, but where a crime has not yet taken place or been proven). It might also create problems with the potential capture, storage, and cross-border sharing of Big Data.

Only three provinces in Canada – Alberta, British Columbia, and Quebec - have their own private sector privacy legislation that supersedes PIPEDA; all others must comply with PIPEDA.

Organizations in Alberta, British Columbia, and Quebec therefore need to be careful of complying with both their own private sector privacy legislation as well as PIPEDA.

Alberta, Saskatchewan, Manitoba, Ontario, New Brunswick, Newfoundland and Labrador and Nova Scotia5have each passed health information protection laws to deal with the collection, use and disclosure of personal health information by public and private sector health care providers. Alberta and British Columbia have also passed privacy laws that apply to employee information. Some of these laws might not be considered to be sufficiently compliant with PIPEDA to be deemed substantially similar. Therefore, in some cases PIPEDA may still apply.

Some provincial sector-specific laws include provisions dealing with the protection of personal information. Most provinces have legislation dealing with consumer credit reporting. These acts typically impose an obligation on credit reporting agencies to ensure the accuracy of the information, place limits on the disclosure of the information and give consumers the right to have access to, and challenge the accuracy of, the information.

Provincial laws governing credit unions typically have provisions dealing with the confidentiality of information relating to members' transactions. There are also a large number of provincial acts that contain confidentiality provisions concerning personal information collected by professionals.

PIPEDA doesn't apply to an organization where it operates entirely within a province that has privacy laws deemed substantially similar to PIPEDA, unless the personal information crosses provincial or national borders.

Each province and territory in Canada is expected to have a commissioner or ombudsman responsible for overseeing provincial and territorial privacy legislation.

(Office of the Privacy Commissioner of Canada)

XI-4: Recent Changes and Other Applicable Privacy Legislation

BILL S-4 THE DIGITAL PRIVACY ACT

Bill S-4 amends the *Personal Information Protection and Electronic Documents Act*,2 the federal private sector privacy law. It does this in several notable ways, including by:

- Permitting the disclosure of an individual's personal information without their knowledge or consent in certain circumstances;
- Requiring organizations to take various measures in cases of data security breaches;
- Creating offences for failure to comply with obligations related to data security breaches; and
- Enabling the Privacy Commissioner, in certain circumstances, to enter into compliance agreements with organizations. (Library of Parliament Research Publications, 2014)

BILL C-13 PROTECTING CANADIANS FROM ONLINE CRIME ACT

Bill C-13 deals with:

- The offence of non-consensual distribution of intimate images;
- Offences committed by means of telecommunication; and
- One aspect of the area of law, generally referred to as "lawful access", an investigative technique used by law enforcement agencies and national security agencies involving intercepting private communications and seizing information where authorized by law.

BILL C-51 INVESTIGATIVE POWERS FOR THE 21ST CENTURY ACT (AKA THE "ANTI-TERRORISM ACT")

Bill C-51 takes into account new communications technologies and equips law enforcement agencies with new investigative tools adapted to computer crimes. The new investigative powers within the legislation give law enforcement agencies the ability to address organized crime and terrorism activities online by:

- Enabling police to identify all network nodes and jurisdictions involved in the transmission of data and the ability to trace the communications back to a suspect. This includes information on the routing, but does not include the content of a private communication;
- Requires a telecommunications service provider to retain data to prevent its loss or deletion while law enforcement agencies obtain a search warrant or production order;
- Makes it illegal to possess a computer virus for the purposes of committing an offence of mischief; and

Enhances international cooperation to help in investigating and prosecuting crimes that extend beyond Canada's borders. (Library of Parliament Research Publications, 2015)

XI-5: AICPA/CICA Privacy Guidelines

The Ten Generally Accepted Privacy Principles

The ten Generally Accepted Privacy Principles are:

1. Management. The entity defines, documents, communicates and assigns accountability for its privacy policies and procedures.

2. Notice. The entity provides notice about its privacy policies and procedures and identifies the purposes for which personal information is collected, used, retained and disclosed.

3. Choice and consent. The entity describes the choices available to the individual and obtains implicit or explicit consent with respect to the collection, use and disclosure of personal information.

4. Collection. The entity collects personal information only for the purposes identified in the notice.

5. Use, retention and disposal. The entity limits the use of personal information to the purposes identified in the notice and for which the individual has provided implicit or explicit consent. The entity retains personal information for only as long as necessary to fulfill the stated purposes or as required by law or regulation and thereafter appropriately disposes of such information.

6. Access. The entity provides individuals with access to their personal information for review and update.

7. Disclosure to third parties. The entity discloses personal information to third parties only for the purposes identified in the notice and with the implicit or explicit consent of the individual.

8. Security for privacy. The entity protects personal information against unauthorized access (both physical and logical).

9. Quality. The entity maintains accurate, complete and relevant personal information for the purposes identified in the notice.

10. Monitoring and enforcement. The entity monitors compliance with its privacy policies and procedures and has procedures to address privacy-related complaints and disputes.

(AICPA/CICA)

XI-6: Other Categories of Personal Information

Sensitive Categories of Personal Information

Some personal information is considered sensitive. Sensitive personal information generally requires an extra level of protection and a higher duty of care. For example, some jurisdictions may require explicit consent rather than implicit consent for the collection and use of sensitive information. Some laws and regulations define the following to be sensitive personal information:

- Information on medical or health conditions
- Financial information
- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Sexual preferences
- Information related to offenses or criminal convictions

Source - (AICPA/CICA)

Non-personal Information

Some information about or related to people cannot be associated with specific individuals. Such information is referred to as *nonpersonal information*. This includes statistical or summarized personal information for which the identity of the individual is unknown or linkage to the individual has been removed. In such cases, the individual's identity cannot be determined from the information that remains, because the information is de-identified or anonymized. Non-personal information ordinarily is not subject to privacy protection because it cannot be linked to an individual. However, some organizations may still have obligations over non-personal information due to other regulations and agreements (for example, clinical research and market research).

XI-7: List of Key Informants

The following key informants were interviewed as part of the research process for this study

- Ms. Marj Akerley
  Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Kelli Brooks
  Principal in Charge, Evidence and Discovery Management
  KPMG LLP
  3020 Old Ranch Parkway, Seal Beach, California, USA
  USA

- Mr. Richard Cumbley
  Partner, Information Management and Data Protection
  Linklaters LLP
  1 Silk Street, London, United Kingdom

- Mr. Howard Deane
  Chair – Emerging Information Technology Committee
  Consumers Council of Canada
  1920 Yonge Street, Toronto, Canada

- Mr. Toundjer Erman
  Director, Business Management Strategic Planning and Business Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Dera J. Nevin
  Director of eDiscovery Services
  Proskauer
  Eleven Times Square, New York, New York, USA

- Mr. Chris Paskach
  Managing Director
  The Claro Group
  350 S. Grand Ave., Los Angeles, California, USA

- Mr. Jean-Sébastien Rochon
  Deputy Director and Counsel,
  National Litigation Support Services/National eDiscovery and Litigation Support Services

- Ms. Dominique Roy
  Director, Business Applications
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Ms. Julie V. Roy
  Supervising Counsel, National Litigation Support Services/National eDiscovery and Litigation Support Services.

- Ms. Tracy Sampson
  Depute Chief Information Officer
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Dugald Topshee
  Director, Client Relationship Management
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Mr. Eric Ward
  Senior Counsel, Public Law Sector – Information law and Privacy Sector
  Canadian Department of Justice
  284 Wellington Street, Ottawa, Canada

- Dr. Anthony Wensley
  Associate Professor, Department of Management,
  University of Toronto Kaneff Centre, 3359 Mississauga RD N Mississauga, Canada

- Mr. Omid Yazdi
  Partner, Forensic Services
  KPMG LLP
  550 South Hope St. Los Angeles, California, USA

Department of Justice, Government of Canada

Big Data Strategic Planning Workshop

November 25, 2015

Outcomes Report

Prepared by E.S. Tunis and Associates Inc.

Table of Contents

1. INTRODUCTION

Senior Department of Justice (JUS) staff members and representatives from different sections of JUS convened on November 25th, 2015 to develop a Big Data strategic plan. The workshop provided them with an opportunity to discuss trends and issues in the application of Big Data in the work of the Department over the next three to five years. Participants worked collaboratively to develop six Strategies that will ensure JUS is well-positioned over the short, medium and long term.

This report describes:

1. The strategic planning framework and the process that was used to develop the six Big Data Strategies.
2. The key Issues that surfaced from research that was conducted on global, national and organizational Big Data trends relevant to JUS over the next three to five years.
3. A Vision for the future of Big Data in the Department of Justice.
4. The six Strategies that will respond to the Big Data Issues and allow JUS to achieve its desired Vision.
5. The Next Steps to implement the Strategies.

Twenty JUS staff and invited guests participated in the event and are listed in Appendix B on page 16. The workshop agenda can be found in Appendix A on page 15.

2. THE STRATEGIC PLANNING PROCESS

The process to develop JUS's Big Data Vision and Strategies had five steps:

Step 1: Context
In order to ground the development of a strategy in empirical evidence, research was undertaken by E.S. Tunis & Associates to identify key external and internal trends that may impact JUS over the next three to five years. (A PowerPoint presentation summarizing the key findings from the research can be found in Appendix D.) The research provided the workshop participants with a platform to launch a discussion of the Issues that the organization is likely to face in the short, medium and long term. The list of Issues generated in this Step formed the basis of the development of the Strategies in Step 3, below.

Step 2: Vision
Workshop participants articulated a Big Data Vision for JUS. The Vision describes how the use of Big Data could "look and feel" in five years, and provides the framework within which JUS can make use of Big Data.

Step 3: Strategy Development
Big Data Strategies are a response to the Issues identified in Step 1 through the lens of the Vision articulated in Step 2. The Strategies are the approach to Big Data that JUS can take over the next three to five years. While the Vision answers *why and what,* the Strategies describes *how* JUS can approach Big Data.

Step 4: Results
Participants defined how success will be recognized should JUS proceed to implement the strategies. Participants identified a variety of results indicators that will help JUS report on progress.

Step 5: Process
In order to implement the Strategies, a significant amount of effort is needed to design plans and processes over the next three to five years. Some steps were taken to identify these processes, but they would need to be developed in full in the coming months.

3. KEY BIG DATA ISSUES

Global, national and internal organizational issues will jointly shape the business environment within which JUS will operate. The Big Data Strategies developed at the workshop are a response to these environmental considerations.

A PowerPoint presentation summarized future Big Data trends. (The presentation can be found in Appendix D.) The participants reviewed and discussed the trends and agreed upon the key Issues that JUS will have to respond to:

1. Governments are often accused of a lack of transparency in the collection and application of Big Data.
2. There is a risk of falling behind global standards and developments in the application of Big Data.
3. There is a lack of adequate human and financial resources dedicated to Big Data in JUS.
4. The impacts of Big Data on individuals (the public) has not been adequately assessed. In addition, there is a deficient legal privacy framework.
5. There are technological, technical architecture, planning and governance barriers to implementing Big Data projects and optimizing organizational productivity.
6. There is organizational and cultural resistance inside and outside JUS to implementing Big Data initiatives.

4. VISION FOR THE USE OF BIG DATA AT JUS

Participants proposed a Vision for the use of Big Data in the organization. The Vision is designed to paint a picture of what Big Data will look like in JUS in the next 3 to 5 years. Participants discussed:

- Why Big Data should be an important part of JUS's work.
- Who Big Data will serve.
- What the Scope of Big Data initiatives will be at JUS.
- What Ethics/Values will be applied to the use of Big Data in JUS.
- What Big Data Governance Structures need to be in place.

WHY JUS Needs to Leverage the use Big Data in its programmes

- Remain current--if JUS is to remain relevant, they need to be on top of Big Data.
- Improve the quality of work at JUS by improving its evidence base.
- Streamline and improve the efficiency of JUS processes.
- Enhance decision making in the organization.
- Make use of data-sets being produced by other organizations.
- Improve JUS's key lines of business.
- Improve public, stakeholder and partner access to JUS.
- Help measure JUS's performance.
- Meet the expectations of the public.
- Anticipate trends.
- Improve data stewardship in JUS.
- Improve public engagement with JUS.
- Liberate the human energy that exists in the organization.
- Improve the accountability and audit functions.

WHO Big Data Initiatives Will Serve

- Canadians--they are JUS's most important client and JUS should be transparent about its use of Big Data.
- Key decision makers in government agencies.
- Legal Services including the Minister of Justice and the Attorney General of Canada, private legal firms and non-governmental organizations.
- Partners including other federal government agencies, provincial territorial and other international governments
- External stakeholders such as academics, Canadian Bar Association.
- Central Agencies such as Treasury Board and the Office of the Auditor General.
- Clients internal to JUS such as Litigation and Legislative Drafting.
- JUS managers. They can be assisted in their programme management by evidenced-based decision making tools supported by Big Data.

The SCOPE of Big Data Initiatives

- All five business lines at JUS are in scope
- Project management.
- Research and innovation.
- Performance and metrics.
- Using Big Data to determine why JUS did not achieve an organizational objective.
- Information governance.
- Using Big Data to contribute to wider discussions in government.
- Using Big Data to facilitate work that is speculative in nature but will have big implications going forward.
- Testing hypotheses to help understand what "we don't know". It can also have a role in identifying what "we don't know we don't know".

The ETHICS AND VALUES that Will Be Applied to the Use of Big Data at JUS

- Treat information and data with the highest ethical standards.
- Uphold the public trust in government.
- Ground Big Data in the idea of wanting to build better public policy.
- Use power responsibly (when one accumulates information you can exercise power).
- The ethic of "do no harm" should be applied to how JUS uses data and information.
- Maintain "open data" values.
- A Big Data view of the world must focus on putting out more data than less.
- Balancing data openness with privacy. (This includes protecting personal identifiers in large datasets.)
- Use model and best practices when dealing with Big Data.
- Full transparency in the application of Big Data.
- Recognize the roles of the Minister of Justice and the Attorney General in the use of Big Data.
- Link Big Data initiatives to the mandate of JUS.
- Use strong stewardship principles in the application of big data.
- Delivering fair and efficient legal operations.
- Ensuring that the use of data is transparent but also intelligible. (A system can be transparent but not intelligible to a lay person.)

Big Data GOVERNANCE at JUS

- There is a need for:
    - a strong policy and legislative framework related to Big Data. From this will flow policies on how to collect data, determine who uses it and how it is used.
    - a governance structure that determines who can have access to what, how data is brought in and how it is used.
    - standards that come from the Big Data governance community.
- Big Data governance at JUS should include a department wide forum on open data.
- A senior management committee should be part of the Big Data Governance structure.
- There are a variety of data sources in JUS (iCase, IFMS, PeopleSoft, shared drives, etc). Each of these have their own governance structure. There should be an overarching Center

for Big Data. This center would cover all of the individual Big Data applications in JUS and have its own governance structure.

- Big Data at JUS could include putting data together in unintended ways. Big Data governance would have to take this into account.
- Big data governance would encompass human resource data.

5. BIG DATA STRATEGIES FOR THE DEPARTMENT OF JUSTICE

The Big Data strategic direction is a response to the Issues, (page 3), through the lens of the Vision (page 4). The proposed five year Big Data strategy could help guide the data activities of the department and allow it to proactively position itself. The Strategies that the participants collectively developed included the following:

- Strategy 1: Improve Data/Information Transparency
- Strategy 2: Become a Big Data "Fast Follower"
- Strategy 3: Effectively Resource Big Data in JUS
- Strategy 4: Protect Privacy
- Strategy 5: Improve the Ability to Access and Use Data in JUS
- Strategy 6: Improve Data Sharing

In order to develop each Strategy, participants undertook the following analysis:

- Problem Definition: Participants clearly defined the problems that each Issue raised and why the problems exist.
- Goal Definition: Participants defined the key objective(s) each Strategy will achieve.
- Results Definition: Participants articulated the measures that will be used to determine if a goal has been met.
- Process Definition: Participants described what actions will be taken to achieve the goal and implement the Strategy[1].

Below is a detailed description of each Strategy and the analysis that the participants used to develop them.

[1] Since the workshop was focused on strategy development and not strategy implementation, only a preliminary discussion of the process for each strategy was undertaken. It is anticipated that further development of the processes will be completed during the operationalization phase of the strategic plan.

Strategy 1: Improve Data/Information Transparency

Problem(s) the Strategy is Addressing:

- There is cultural resistance to data transparency within JUS. There is a lot of good information that can be shared with the public but there is internal resistance to "put it out there".
- There is active resistance to sharing information within JUS.
- When it comes to data sharing, the default mode at JUS is "secrecy".
- There are perceived risks to being transparent with government data and there is a perceived lack of "permission" to share data.
- It takes time and money to effectively share information. A lot of effort is required to prepare information before it can be released.
- It is not clear what kind of information JUS should protect and what should be released.
- Some staff fear that if certain information were freely available, they could be out of a job.
- Some data in JUS is not anonymous.
- There is a cultural fear of negative exposure within JUS. I.e. there is concern about what the public will say/think when they learn more about how government works.
- Some of the key reasons the above problems exist is because:
    - It is not a natural reflex for JUS to be transparent with data.
    - There is a culture of risk aversion and a fear of error within JUS.

The Goal(s) of the Strategy:

The participants want Canadians to have increased access to justice data and be engaged and informed about the workings of JUS and the justice system.

Results that Indicate if the Goal of the Strategy Has Been Achieved:

- Canadians will have quality information for which they can hold government to account.
- People outside the department will have better understanding of the business of JUS.
- JUS resources will be utilized more effectively. (E.g. Less time will be spent on activities such as dealing with ATIP requests.)
- There will be higher quality, more intelligent and more democratic discussions about the Canadian justice system.

Preliminary List of Processes to Achieve the Goal:

- Share information with the public.
- Change the culture in JUS so that it is not averse to data transparency and information sharing.
- Determine a schedule for what and when data should be released.
- Engage the public through initiatives like public legal education sessions.

Strategy 2: Become a Big Data "Fast Follower"

Problem(s) the Strategy is Addressing:

- Other governments and private sector legal firms are quickly increasing their usage of Big Data.
- JUS is, currently, not in a position to take advantage of Big Data, whether it is internal or external data.
- JUS will be in comparative disadvantage if it is not current with Big Data techniques and technology.
- JUS's ability to provide good advice to its clients and the government will be compromised if it does not make effective use of Big Data.
- Other organizations and departments in the government of Canada have Big Data strategies--JUS should not fall behind them.
- The evidence the department uses to support policy development and litigation will be weak if JUS does not keep pace in the development of Big Data.
- Big data will be key to acquiring good evidence. Not having good evidence affects how well policy and programmes do. If JUS does not have good evidence, then a policy or programme can be expected to fail.
- One of the key reasons the above problems exist is because JUS is in a reactive mode and Big Data requires that organizations be proactive.

The Goal(s) of the Strategy:

The Department of Justice could investigate and pursue technologies and techniques that have been proven to work effectively. I.e. JUS will be a "fast follower" since JUS primary business in not data collecting.

Results that Indicate if the Goal of the Strategy Has Been Achieved:

- JUS will be well positioned, with respect to Big Data, within the government of Canada.
- JUS will be positioned to take advantage of large data-sets.
- JUS will play a new federal role—a Big Data role—in the justice system.

Preliminary List of Processes to Achieve the Goal:

- JUS ~~to~~ could be an active follower, with respect to Big Data, in the justice community.
- Take advantage of lessons learned by other jurisdictions.
- Collaborate, facilitate and share information with internal and external organizations/departments that have successful Big Data programmes.

Strategy 3: Effectively Resourcing Big Data in JUS

Problem(s) the Strategy is Addressing:

- Adequate resources (human and financial) have not been targeted to Big Data in JUS, more of a reallocation of some current resources could be sufficient for the purposes of JUS.
- There is no ownership of Big Data in JUS. Staff tend to say "it is not my job or not my problem to fix".
- There is a lack of an overarching coordination mechanism that will allow JUS to take full advantage of Big Data.
- There is not enough Big Data technical expertise in JUS.
- The Big Data human resources and financial resource deficiency is not seen as a problem in JUS, particularly since the level of activity would not be extensive. Effective human and financial resourcing of Big Data is not undertaken in JUS.

The Goal(s) of the Strategy:

A departmental commitment to effectively resource Big Data initiatives with demonstrable results.

Results that Indicate if the Goal of the Strategy Has Been Achieved:

- Buy-in at the departmental level with respect to the idea that Big Data is an important part of JUS's work.
- A single point of contact for Big Data in JUS could be established.
- Adequate human and financial resources invested in Big Data.
- High visibility of Big Data in JUS. What Big Data is and how it is applied will be clearly understood by staff in the department.
- The value for money (ROI) of Big Data will be clear and provable.

Preliminary List of Suggested Processes to Achieve the Goal(s):

- Identify a Big Data champion.
- Initiate kick-start projects for Big Data in JUS.
- Initiate Big Data pilot projects.
- Ensure Big Data projects are of sufficient priority.
- Develop a business case for investing human and financial resources into Big Data.

Strategy 4: Protect Privacy

Problem(s) the Strategy is Addressing:

- There is a risk of sensitive information being released to the public. (Especially as it pertains to marginalized groups.)
- There is a risk of privacy laws becoming out of date.
- The general public has a distrust of government and a distrust in its ability to manage and protect information.
- There is a fear that personal information can be reconstructed from multiple data sources. (I.e. A researcher may be able to take two data sources, combine them and identify who a person is.)
- People do not have a clear understanding of how much information is being collected.
- The younger generation is not concerned enough about how their personal information is collected and used.

The Goal(s) of the Strategy:

Ensure personal information is protected when Big Data tools and techniques are used at JUS. This includes respecting the privacy of the public and ensuring JUS is a trustworthy and responsible user of Big Data.

Results to Measure if the Goal of the Strategy Has Been Achieved:

- An increased level of trust in JUS.
- The values of JUS will be "embedded" in Big Data technology. This would include ensuring IT development standards and Big Data initiatives respect personal privacy.
- There will be public engagement in the various JUS tools. (I.e. People will be willing to participate in JUS social media tools, apps, etc.)

Preliminary List of Processes to Achieve the Goal:

- Create a framework that will protect personal information when Big Data initiatives are undertaken.
- Improve the information security framework especially as it pertains to Big Data initiatives.
- Establish IT development standards that will ensure personal data privacy is maintained.

Strategy 5: Improve the Ability to Access and Use Data in JUS

Problem(s) the Strategy is Addressing:

- JUS is not optimizing its use of the data tools it already has in-house.
- The data tools that JUS has do not "talk" to each other.
- Staff do not want to use the data tools that are provided to them. (Note: This is a cultural issue within the organization.)
- Some of the data tools in JUS do not do what staff want them to do.
- There are barriers to accessing data in JUS.
- There are overlapping and duplicated data-sets in JUS.
- It is difficult to access data and data repositories that are valuable and can be used for Big Data initiatives.
- There is no clear understanding of the Big Data "landscape" in JUS.
- Staff do not have direct access to a lot of their own data in JUS.
- There are data sources (e.g. legal services) that are collated in many different departments.
- Some of the data-sets are of low quality in JUS. This is because data is often not entered properly or not entered at all.
- A lot of the data in JUS is decentralized. It is not clear where all of the data in the department is stored.
- It is not clear how much of the data in JUS is useful.
- Some of the key reasons the above problems exist are because:
    - There is decentralized ownership of data in JUS.
    - There is diffused accountability for data collection and management.
    - There is a lack of formal data standards.
    - Data management has never been a priority in JUS.
    - There was never a move to centralize data because of lack of awareness that Big Data would become an issue.
    - There are challenges with document management, information management and data input in JUS which it is now attempting to improve (a good time to include a Big Data lens to that work).

The Goal of the Strategy:

Data in JUS will be easy to find, access, retrieve, use and analyse.

Results to Measure if the Goal of the Strategy Has Been Achieved:

- Staff will be able to find the data and information they are looking for in JUS and they will be able to do it quickly.
- Big Data tools will be used on a regular basis.
- There will be no bottlenecks that restrict access to data.

Strategy 6: Improve Data Sharing

Problem(s) the Strategy is Addressing:

- Big data threatens the "power" of current information holders in JUS.
- There is a tradition of closed data in JUS.
- Staff may resist the conclusions that come from the analysis of Big Data.
- Staff may be concerned about how Big Data would be used to evaluate performance. (Some staff feel they have been "burned" by performance metrics in the past.)
- There is lack of trust in data in the organization.
- Some staff believe sharing information will require too much effort.
- Some staff are concerned that if their data is taken from them they will lose control or they will be judged. (They feel there is personal risk in sharing their data.)
- There may be external resistance to JUS consolidating and integrating data.

The Goal(s) of the Strategy:

Data sharing should be the default mode in JUS.

Results to Measure if the Goal of the Strategy Has Been Achieved:

- The public, stakeholders, special interest groups, lobby groups, academics and members of the legal profession will be engaged with JUS.
- The public will feel JUS is relevant.
- The legal community will be engaged with Big Data initiatives that JUS undertakes.

Preliminary List of Processes to Achieve the Goal:

- Ensure staff learn how their jobs can be enhanced by performance analytics.
- Get senior management onboard with Big Data. Show them how it works and what it can do.
- Work to ensure that staff who are culturally resistant to sharing data are onboard.
- Create tools that make sharing data easy. (This includes demonstrating that sharing information will not create additional work.)
- Communicate to staff that sharing data will be valuable to them.
- Develop a change management plan for Big Data. (This includes identifying the groups that will need the most cultural change.)
- Use Big Data to interact and build confidence with the public.

6. NEXT STEPS

Participants proposed the following Steps to move forward with the Big Data Strategy process.

- An Outcomes Report from the Big Data workshop will be created that contains JUS's Vision and six Strategies for Big Data that could be implemented over the next 3 to 5 years. The Report will be shared with all of the workshop participants and they will be invited to provide feedback. The feedback will be integrated into the final Report. The finalized version will be shared with all staff in the organization, starting with the Deputy Minister and the Chief Information Officer. The Report will also be shared with and "talked-up" at The Business Transformation Committee, The Policy Committee, by colleagues at Public Safety and colleagues at Statistics Canada.

- A proposal for a proof of concept/business case for implementing the Strategic Plan could be developed.

- Finally, following the workshop, participants have committed to sharing the below messages with colleagues regarding the event:
    - The participants learned something new about analytics and about how they can be applied at JUS.
    - New Strategies about how to approach Big Data were developed.
    - In the coming months, the participants will continue building on the work they started at the workshop.
    - The participants will build partnerships to keep the strategic planning process moving.
    - The work that was completed at the workshop will be moved forward to the Deputy Minister.
    - The workshop was a good example of horizontal collaboration.

APPENDIX A: WORKSHOP AGENDA

Justice Canada
Big Data Workshop

Wednesday, November 25, 2015

Location: Library & Archives Building, 395 Wellington Street, Room 156

Objective: The objective of the workshop is to consider the research on forward trends and associated issues in the use of Big Data in the JUS legal environment and to consider what a Big Data Strategy for the department could be.

AGENDA

8:30 to 8:40	**Welcome and Opening Remarks** Stan Lipinski, Director General, Policy, Integration, and Coordination Section, Policy Sector
8:40 to 9:10	**Introduction to the Agenda and the Strategic Planning Framework** Workshop facilitators from E.S. Tunis and Associates (ESTA) will review the strategic planning process and framework being used for this workshop.
9:10 to 11:30 (With break at 10:00)	**Big Data Research – What Issues does Big Data pose for JUS?** A research paper commissioned by the Research and Statistics Division has examined the possible uses of Big Data in legal environments. It has identified some future trends and related issues for JUS to consider. Each set of trends and issues will be discussed and debated by participants.
11:30 to 12:30	**A Vision for Big Data in Justice Canada** Participants will be asked to construct a forward Vision for the positioning of Big Data in Justice Canada.
12:00-12:30	Working Lunch- (Lunch will be served in the room)
12:30 to 1:30	**Building a Big Data Strategy**

	Participants will be introduced to a model of strategy building and will address the first set of issues from the Big Data Research
1:30 to 3:45	**Continuing to build the Big Data Strategy**
(With break at 2:30)	Participants will break into small groups of their choice to articulate strategic responses to the remaining issues from the Big Data Research
3:45-4:30	**Big Data Strategy Review**
	Participants will present their strategies to the plenary session
4:30 to 5:00	**Next Steps**
	The facilitators will summarize the products of the day and outline next steps in the development of the Big Data Strategy for Justice Canada

APPENDIX B: WORKSHOP PARTICIPANT LIST

(The invitation list was wider to include areas that may have a role or interest in Big Data, however, a number of invited participants could not attend)

Name	Organization
Dugald Toshee	Information Solution Branch, Management and CFO Sector
Natasha D'Souza	Human Resource Branch, Management and CFO Sector, Justice Canada
Eric Ward	Information Law and Privacy Section, Public Law Sector
Susan Fisher-Clement	Communication Branch
Charlotte Fraser	Research and Statistics Division, Policy Sector
Jacque Ouellette	Information Solution Branch, Management and CFO Sector
Stan Lipinski	Policy, Integration, and Coordination Section, Policy Sector
Lynn Barr-Telford	Health, Justice, and Special Surveys, Canadian Centre for Justice Statistics, Statistics Canada
Paul Roy	Senior Assistant Deputy Minister's Office, Policy Sector
Minelle D'Souza	Library Services, Management and CFO Sector
Jean-Sebastien Rochon	National eDiscovery and Litigation Support Services, Litigation Branch
Ryan Hum	Central Innovation Hub, Privy Council Office
Ting Li	Research and Statistics Division, Policy Sector
Mala Khanna	Information Law and Privacy Section, Public Law Sector
Andrew Fobert	International Assistance Group, Litigation Branch
Bill Bedford	Finance and Planning Branch, Management and CFO Sector
Michel Champagne	Communication Branch
Claudie Besner	Finance and Planning Branch, Management and CFO Sector
Peter Beaman	Legislation and Regulations Group, Legislative Services Branch
Esther Rubenstein	Strategic Policy and Research Division, Public Safety Canada

APPENDIX C: WORKSHOP EVALUATION

At the end of the workshop, the participants were invited to provide feedback on the day's activities by indicating their agreement or disagreement with the following statements:

- "The Big Data Strategy workshop was a good use of my time."
- "I have a better understanding of Big Data after having participated in the workshop."
- "I am willing to contribute to this discussion again."

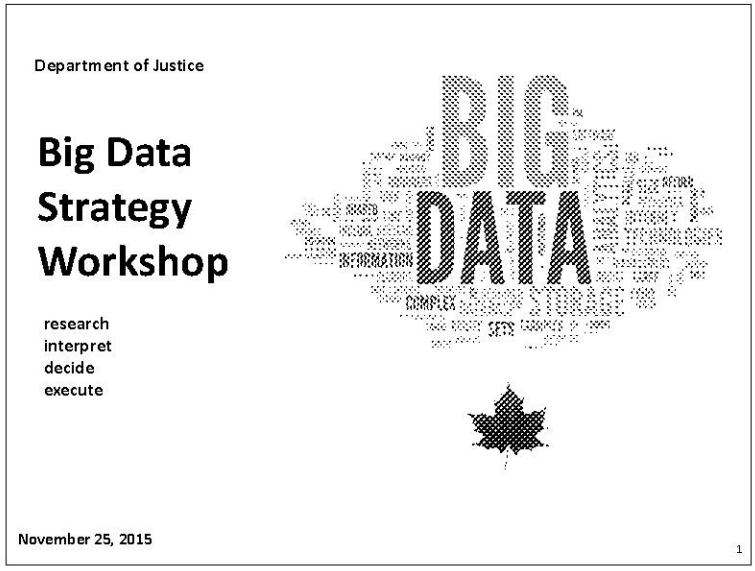
The below table contains the results of the evaluation.

	Participant Responses	
	Yes	No
The Big Data Strategy workshop was a good use of my time.	9	0
I have a better understanding of Big Data after having participated in the workshop.	8	1
I am willing to contribute to this discussion again.	9	0

APPENDIX D: BIG DATA TRENDS RESEARCH

The below PowerPoint presentation contains the key global, national and organizational Big Data trends that may be relevant to JUS over the next three to five years. It also contains a set of possible issues that the trends raise for the Department.

Note: The presentation was developed based on the research report conducted by E.S. Tunis & Associates, Inc. titled "Possible Big Data Uses by the Department of Justice And Related Privacy Concerns", September 2015.



DEPARTMENT OF JUSTICE BIG DATA STRATEGY WORKSHOP

WORKSHOP OBJECTIVES AND AGENDA

Objectives

- Consider the research on forward trends and issues in the use of Big Data in the Department of Justice legal environment.
- Develop a Big Data Strategy for the Department of Justice.

Agenda

- Morning [8:30 to 12:00]
  1. Welcome and Opening Remarks
  2. Overview of the Strategic Planning Framework and Process
  3. Big Data Research—What issues does Big Data pose for JUS?
  4. A Vision for Big Data in the Department of Justice.
- Working Lunch [12:00 to 12:30]
- Afternoon [12:30 to 5:00]
  5. Building a Big Data Strategy (Small Group Work)
  6. Big Data Strategy Review (Small Groups Report to Plenary)
  7. Next Steps

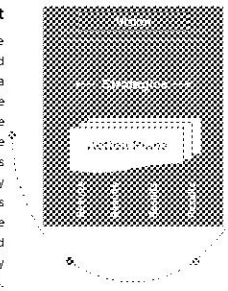
2

DEPARTMENT OF JUSTICE BIG DATA STRATEGY WORKSHOP

THE STRATEGY FRAMEWORK AND PROCESS

Step 1: Context

Strategy development should always be grounded in evidence-based data and trends which explore various Big Data opportunities and challenges that the Department of Justice will face in the short, medium and long terms. These are captured in the research findings and will be discussed and debated by the workshop participants. This process will generate a list of issues which are likely to be faced in the future and which will form the basis for strategy development in Step 3.



Step 4: Results

Truly successful organizations don't stop the strategy process after the development of a vision and strategies. The critical next step is the clear articulation of expected results so that leaders can measure the success, or failure, of the organization in working toward their vision. Each strategy will be developed to include a variety of indicators that will help the Department report on its progress and make course corrections as necessary.

Step 2: Vision

The vision statement provides the framework within which the Department of Justice will operate over the next three years with respect to Big Data. A vision statement describes ***how*** Big Data will "look" in the Department.

Step 3: Strategy

The strategy is the response to the issues identified in Step 1 through the lens of the vision articulated in Step 2. Strategy is the approach the Department will take over the next three years with respect to Big Data.

BIG DATA RESEARCH FINDINGS

01 Research Methodology

02 Global Trends

03 Trends in Canada

04 Trends in Justice Canada

4

1

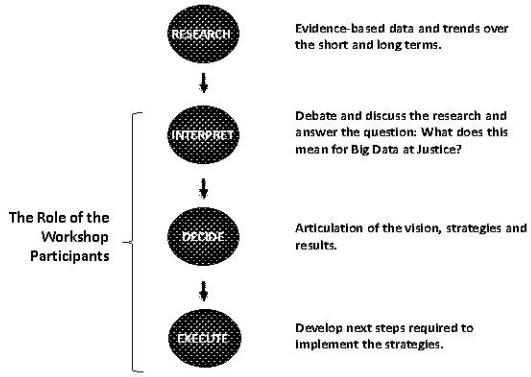
Research Methodology

- Overview of the Strategy Development Process
- Research Data Sources

5

1. METHODOLOGY

OVERVIEW OF THE STRATEGY DEVELOPMENT PROCESS



6

1. METHODOLOGY

DATA SOURCES

- Literature review.
  - Published Reports
  - Research Papers
  - Web Articles
  - Media Reports
- Interviews with key informants.
  - Internal: Department of Justice staff and managers.
  - External: Academics, law firms, providers of big data services, private sector users of big data services and non-governmental organizations.

7

2

Global Trends

- Global Big Data Trends in Government Institutions
- Global Big Data Trends in Legal Systems

8

TABLE OF CONTENTS

9

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN GOVERNMENT INSTITUTIONS

Increase in Information Sharing Across Government Agencies

- There is an increase in the sharing of information across government agencies.
  - Federal governments recognize the value of having access to data across agencies.
  - Sharing of information will create complex inter-departmental relationships that will make transparency more difficult.
  - There will also be an increase in the need for staff skilled in predictive coding and structured data analytics in order to analyze complex cross-agency information systems.

10

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN GOVERNMENT INSTITUTIONS

Increased Availability of Internet Listening Tools

- Software solutions and Internet "listening tools" that can monitor and track multiple social media channels and analyze trends, including public sentiment, are becoming increasingly available to governments.
  - Complex systems and complex data-sets will be required for these tools to impact government decision-making.

Data Privacy Legislation Will Continue to Evolve

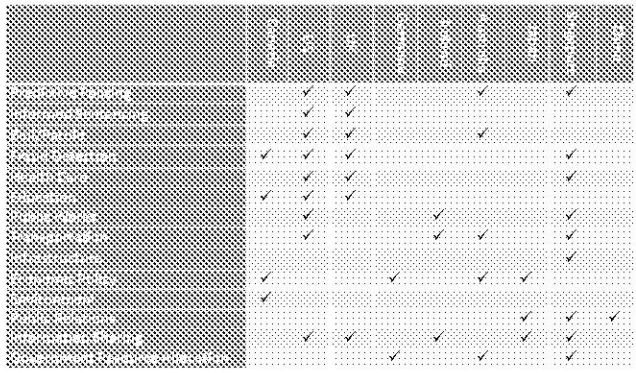
- All governments will continue to struggle with ways to keep their data privacy legislation current, relevant, and usable as Big Data technology rapidly develops.
  - Data mining tools and the consolidation of departmental databases will have a particularly significant influence on legislation.

11

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN GOVERNMENT INSTITUTIONS

Use of Big Data in Other Federal Governments



12

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN LEGAL SYSTEMS

eDiscovery Tools Will Become More Prevalent in Legal Organizations

- eDiscovery tools will continue to evolve as the types of electronic evidence expand from email, documents and voice mail to include social media and mobile data.
- In 2013, the global eDiscovery market was estimated to be USD $5.56 billion. Government agencies were the largest end-user segment, accounting for 51% of the revenue share. The market is expected to grow at an annual rate of 15.5% over the next five years.
- For legal organizations to capitalize on the use of Big Data for eDiscovery it will be necessary for them to have:
  - An appropriate information management governance structure.
  - A limited set of standardized eDiscovery tools that permit legal counsel to become experienced with their use.

13

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN LEGAL SYSTEMS

Increase in the Use of Big Data for Case Assessment and Settlement

- The use of Big Data for case settlement and dispute resolution processes is expected to be one of the most significant future uses of Big Data in the judicial system.
- Many large legal firms are adopting the use of early case assessment software to estimate the legal risk of prosecuting or defending a case based on the financial resources required.

Increase in the Ability to Predict the Outcome of Cases

- The combination of Big Data and artificial intelligence tools will increasingly allow users to predict the outcome of cases.
  - New statistical correlations are being discovered using historical legal case datasets.

14

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN LEGAL SYSTEMS

Predicting Case Outcomes: Examples

Guilty Verdict Probabilities: Selected Examples

	Impaired Driving	18 to 24 Years	Male	93.8%
	Theft	25 to 34 Years	Male	89.5%
	Fraud	18 to 24 Years	Female	83.8%
	Common Assault	55 Years and Over	Male	36.1%
	Possession of Stolen Property	18 to 24 Years	Female	20%

Predicted Case Length Estimates: Selected Examples

	Sexual Assault	35 to 44 Years	Male	531
	Robbery	18 to 24 Years	Male	247
	Crimes Against Property	25 to 34 Years	Female	182
	Theft	18 to 24 Years	Female	23
	Impaired Driving	18 to 24 Years	Male	1

Source of Dataset: Over 3 million adult criminal court cases from Statistics Canada. 15

2. GLOBAL TRENDS

GLOBAL BIG DATA TRENDS IN LEGAL SYSTEMS

Privacy Laws in the International Community Will Impact Canada

- Privacy laws in the international community are far from static, and changes are likely to have an impact on Canadian laws and practices as they evolve.

16

2. GLOBAL TRENDS

ISSUES RELEVENT TO JUSTICE—FOR DISCUSSION

Issue 1: Falling Behind Global Standards

- The legal world seems headed down a path where sophisticated Big Data and analytics technologies will play an increasing role in managing litigation costs and achieving success in the trial process. Is Justice Canada falling behind global standards? If so, will it be at a comparative disadvantage?
- Are there any activities that have been noted in other countries or government departments that might be useful for consideration by Justice Canada?

17

2. GLOBAL TRENDS

ISSUES RELEVENT TO JUSTICE—FOR DISCUSSION

Issue 2: Lack of Adequate Human and Financial Resources

- Significant investment will likely be required over many years, in financial and human resources to remain current with the external legal marketplace, especially if there are significant developments in the use of Big Data and predictive analytics technology:
  a. Justice Canada may not have access to the financial, human, and other resources required to move down all the emerging technology paths at once.
  b. How can resource needs be best prioritized as part of an overall departmental strategy?
  c. How might the ongoing commitment to invest in the changes and tools required be ensured over a protracted period of time?

18

3

Trends in Canada

- Big Data Trends in Canada
- Big Data Privacy in Canada

19

TABLE OF CONTENTS

2
0

3. TRENDS IN CANADA

BIG DATA TRENDS IN CANADA

Increasing Availability of Government and Legal Information to Canadians and Lawyers

- An increasing amount of information is available from the Canadian Government through "Open Government" and other initiatives--this trend is likely to continue.
- The Canadian Legal Information Institute (CanLII), a non-profit organization managed by the Federation of Law Societies of Canada, has a free legal database that is rapidly becoming one of the tools of choice for legal research.
  - CanLII provides lawyers with access to court judgments, tribunal decisions, statutes and regulations from all Canadian jurisdictions.

21

3. TRENDS IN CANADA

BIG DATA AND PRIVACY IN CANADA

Canadians Increasingly Concerned About Privacy

- As sharing of Big Data by government agencies becomes more commonplace, it is expected that Canadians will become increasingly concerned about privacy. According to a study conducted by the Office of the Privacy Commissioner (OPC) in December 2014:
  - 9 in 10 Canadians expressed some level of concern about the protection of their privacy.
  - 34% of Canadians indicated that they were ***extremely concerned*** about their privacy.
  - 73% indicated they have less privacy than they did 10 years ago.
  - 78% indicated that they were either very or somewhat concerned about government surveillance.

22

3. TRENDS IN CANADA

BIG DATA AND PRIVACY IN CANADA

Increasing Public Debate About Data Sharing Between Departments

- The Privacy Commissioner of Canada's comments on Bill C-51 and his position on information sharing is likely to create further debate and shape public opinion regarding government Big Data and data sharing between departments and with other governments.

  "In my view, Bill C-51, in its current form, would fail to provide Canadians with what they want and expect: legislation that protects both their safety and their privacy.... The scale of information-sharing between government departments and agencies proposed in this bill is unprecedented. The new powers that would be created are excessive and the privacy safeguards proposed are seriously deficient."

23

3. TRENDS IN CANADA

BIG DATA AND PRIVACY IN CANADA

Greater Emphasis on Public Disclosure to Reduce Privacy Concerns

- It is expected that in order to address the privacy concerns of Canadians with respect to how their personal information is being used, greater emphasis on public disclosure will become more important to government agencies.

Information Sharing Will Increasingly Raise Jurisdiction Issues

- Determining which jurisdiction governs personal information is becoming much more complicated as information is gathered and/or transferred across legal jurisdictions and co-mingled in Big Data stores.

24

3. TRENDS IN CANADA

ISSUES RELEVENT TO JUSTICE—FOR DISCUSSION

Issue 3: Risk of Impact of Big Data on Individuals Not Adequately Assessed

- How should the risk to individuals be assessed, together with the cost and effectiveness of putting mitigating controls in place as a part of the business case for implementing Big Data solutions? What should be the main considerations by Justice Canada?

Issue 4: Insufficient Government Transparency

- Should the implementation of a Big Data repository by government require greater government transparency about the way in which government handles personal information in Canada? What factors need to be considered by Justice Canada?

Issue 5: Lack of Adequate Privacy Legal Framework

- Are the various laws affecting personal data privacy in Canada adequate to deal with the emerging uses of data, and especially as government and other Big Data projects are brought on stream?

25

4

Big Data Trends in the Department of Justice

- Future Use of Big Data in Justice Canada
- Big Data Tools in Justice Canada

26

TABLE OF CONTENTS

27

FUTURE USE OF BIG DATA IN JUSTICE CANADA

Primary Applications of Big Data

- The primary applications of Big Data analysis in the Department of Justice are expected to be the use of:
  - eDiscovery software tools for analysing evidence for trials.
  - Predictive analytics to predict the outcome of cases.
  - Data analytics techniques to analyze operational statistics to improve JUS productivity and cost-efficiency.
  - Data analytics to predict environmental trends for use in guiding changes in government policies.
  - Tools for early management of legal risks associated with individual cases.
  - Tools to measure performance and compliance with departmental and professional policies, procedures and standards.

28

FUTURE USE OF BIG DATA IN JUSTICE CANADA

Involvement in Legal Actions Surrounding Personal Data

- Justice Canada appears unlikely to capture and use a significant amount of personal Big Data.
  - Regardless, Justice Canada is likely to be involved in legal actions or discussions surrounding the use of personal data by other government departments.
  - In such cases, Justice Canada lawyers will need to respect their obligations under the Personal Information Protection and Electronic Documents Act (PIPEDA).

29

4. BIG DATA TRENDS IN THE DEPARTMENT OF JUSTICE

BIG DATA TOOLS IN JUSTICE CANADA

Increase in Interest in Advanced Big Data Software Tools

- The Justice Canada IT Department is increasingly interested in Big Data analysis software. E.g.
  - HP's Autonomy: Allows for the analysis of large scale unstructured Big Data repositories.
  - ROSS, an experimental artificial intelligence system built on IBM's "Watson" artificial intelligence platform developed by researchers at the University of Toronto.

Restructuring of Justice Canada's Core Evidence Management Tool

- Plans are in place to restructure the core evidence management tool ("Ringtail") so that over 25 million pages of documents across the system can be searched.

30

4. BIG DATA TRENDS IN THE DEPARTMENT OF JUSTICE

BIG DATA TOOLS IN JUSTICE CANADA

Increased Usage and Improvement of Microsoft SharePoint

- Steps are being taken to migrate litigation documents in the iCase tool to Microsoft SharePoint 2013.
- Justice Canada is looking to incorporate Fast Search capability into Microsoft SharePoint 2013 to improve the performance and cross-government Big Data search capability of its content search engine.

31

4. BIG DATA TRENDS IN THE DEPARTMENT OF JUSTICE

ISSUES RELEVANT TO JUSTICE—FOR DISCUSSION

Issue 6: Insufficient Use of Big Data Tools to Optimize Organizational Productivity

- In order to enhance organizational productivity, should the Department be more focused on:
  a) Tools that improve research and evidence gathering?
  b) Using Big Data to improve departmental administration and operations?
  c) Using analytics to allocate litigation resources more effectively by predicting future demand and workload?
  d) Using Big Data to reduce litigation costs by predicting case outcomes and resolving cases without going to trial?
  e) Using access to internal and external Big Data sources to permit better policy decisions?

32

4. BIG DATA TRENDS IN THE DEPARTMENT OF JUSTICE

ISSUES RELEVANT TO JUSTICE—FOR DISCUSSION

Issue 7: Technological Barriers to Implementing Big Data Projects

- Are there any serious technology impediments that would prevent the department from moving forward with Big Data projects?

Issue 8: Planning Barriers to Implementing Big Data Projects

- The implementation and adoption of complex new technologies is a significant undertaking and will take considerable time to accomplish. What advance planning and action will be required to ensure that the required resources are available to make implementation a success?

Issue 9: Organizational Resistance to Implementing Big Data Initiatives

- What resistance can be expected to the changes required, and what sort of change management program might be required to overcome resistance?

33

DEVELOPING A BIG DATA STRATEGY FOR THE DEPARTMENT OF JUSTICE

RECAP OF THE ISSUES

Issue 1: Falling Behind Global Standards

Issue 2: Lack of Adequate Human and Financial Resources

Issue 3: Risk of Impact of Big Data on Individuals Not Adequately Assessed

Issue 4: Insufficient Government Transparency

Issue 5: Lack of Adequate Privacy Legal Framework

Issue 6: Insufficient Use of Big Data Tools to Optimize Organizational Productivity

Issue 7: Technological Barriers to Implementing Big Data Projects

Issue 8: Planning Barriers to Implementing Big Data Projects

Issue 9: Organizational Resistance to Implementing Big Data Initiatives

34